Barcelona Supercomputing Center
Centro Nacional de Supercomputación

# Data visualization for researchers

## Crash course

Guillermo Marin

# What is visualisation?

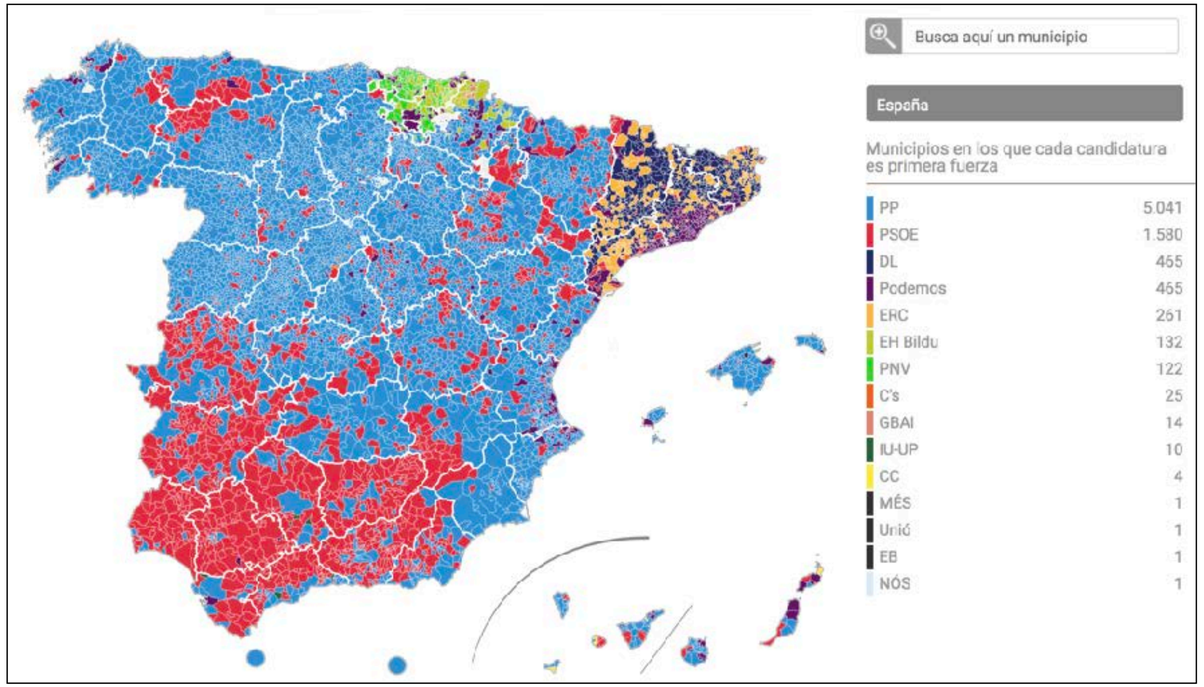
# The visual representation of complex information in ways that lead to insight

## Limited capacity to extract information from abstract representations

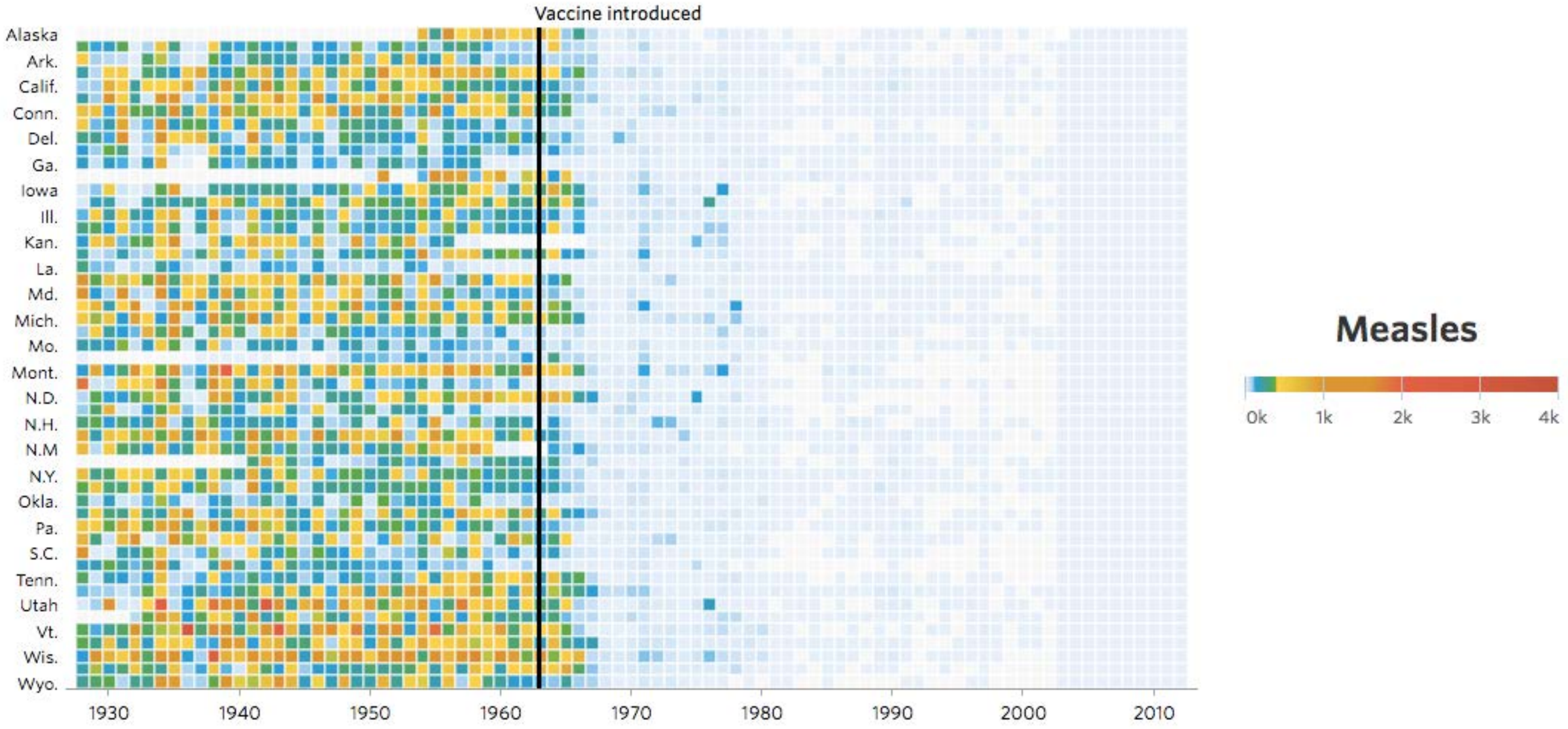
| | A | B | C | D |
|---|---|---|---|---|
| 1 | Libro | Autor | Donde esta | Tema |
| 27 | Trees, maps, and theorems. Effective communication for rational minds | Jean-luc Doumont | Armario | Visualization |
| 28 | International yearbook communication Design 2016/2017 -Vol 1 | red dot edition | Armario | |
| 29 | International yearbook communication Design 2016/2017 -Vol 2 | red dot edition | Armario | |
| 30 | Practical Data Science with R | Nina Zumel & John Mount | Armario | Data Science |
| 31 | Python for Data Analysis | | Irene | Data Science |
| 32 | Advanced Analytics with Spark | Sandy Ryza, Uri Laserson, Sean Owen & Josh Wills | Carlos | Data Science |
| 33 | Learning Spark | Holden Karau, Andy Konwinski, Patrick Wendell & Matei Zaharia | Carlos | Data Science |
| 34 | Visual Strategies | Felice C. Frankel & Angela H. DePace | Armario | Visualization |
| 35 | Visual Thinking for Design | Colin Ware | Guille | Visualization |
| 36 | Building Responsive Data Visualization for the web | Bill Hinderman | Armario | Visualization |
| 37 | Dear Data | Giorgia Lupi & Stefanie Posavic | Guille | Visualization |
| 38 | Introducción al diseño de información | Kathryn Coates, Andy Ellison | Luz | Visualization |
| 39 | La información en el diseño | Isabel Meirelles | Guille | Visualization |
| 40 | Raw Data | | Armario | Visualization |
| 41 | GOETHE Teoría de los colores | Wolfgang von Goethe | Armario | Visualization |
| 42 | The Truthful Art (ejemplar 1) | Alberto Cairo | Armario | Visualization |
| 43 | Information Made Beautiful - infographic design | Sandpoints | Guille | UI Design |
| 44 | Data Points | Nathan Yau | Fernando | Visualization |
| 45 | Information Visualization | Colin Ware | Armario | Visualization |
| 46 | Javascript: The Good Parts | Douglas Crockford | Fernando | Programming |
| 47 | Learning Python | Mark Lutz | Guille | Programming |
| 48 | Save the cat! | Blake Snyder | Armario | Writing |
| 49 | Statistics Done Wrong | Alex Reinhart | Carlos | Data Science |

```
3.892255132300072584e-01 4.440158073305188191e-01 -1.960414963285537493e-01 -2.682237493715293741e-01 1
5.145199279912325563e-02 -1.972227878769014231e-01 5.424776248819125346e-01 5.968651360223455049e-02 7.
-1.773566561022106303e-01 -9.220018292576756189e-03 3.855770202771902633e-01 5.968416363905423294e-01 1
2.553912064985732755e-01 3.897469299728302250e-01 -1.194835528549778592e-02 1.785980443778885374e-01 3.
-2.627083935891604272e-01 -1.822086594180704344e-01 -4.389616887225274988e-01 -3.967905143824700387e-01
7.188871767116146450e-02 2.992344255415894660e-01 -3.028602544511632133e-01 -3.321291871957253172e-01 -
4.669270134157175178e-01 1.689636044906402679e-01 -8.123112057242825104e-03 4.937582253267484012e-01 3.
-3.819915668876001091e-01 -3.309379706489690848e-02 -5.166199568847842233e-02 2.450274748290849069e-01
5.559055273379925194e-02 5.579410785291329944e-01 -5.437982492087630737e-01 -9.711848550036226990e-02 -
2.173679879407889037e-01 -4.303627691564007596e-01 1.777409344888698089e-02 -4.297273956670333295e-01 3
3.693889914441865208e-01 1.091509207325834477e-01 -7.681560537473072769e-02 -2.213424167159359357e-01 -
2.452716007539893495e-01 4.019921062998677019e-01 -3.376145187243325685e-01 -3.088943380345906542e-01 -
-2.257191674889425514e-01 -2.154388730285226872e-01 9.055377721261199941e-02 -5.404528937196890342e-02
2.036153995195968125e-01 2.053859544767734358e-01 2.779921657253315304e-01 -1.629364457946377021e-01 2.
1.794606477301371183e-01 3.373411329112541157e-01 2.796311659865462484e-01 7.600671710970950956e-01 3.
-3.487288363015947357e-01 -2.731762068145069874e-03 0.000000000000000000e+00 5.163047591401367237e-02 3
```

## Cognitive mechanisms specific for patern detection




Elecciones generales 2016 - Pablo Medina, El Mundo



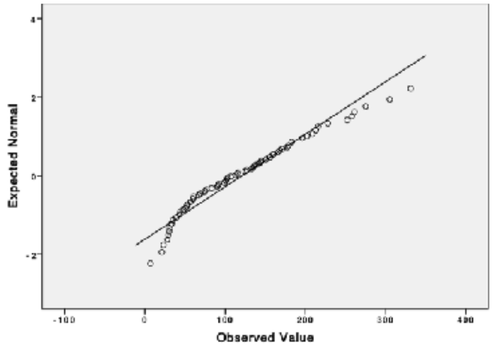

http://graphics.wsj.com/infectious-diseases-and-vaccines/

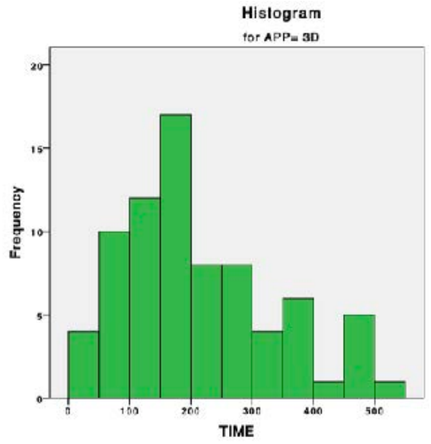
# Why visualise at all?

## Exploration

### Explore the data / Raise questions

Bar charts
Pie charts
Scatter plots
Traces



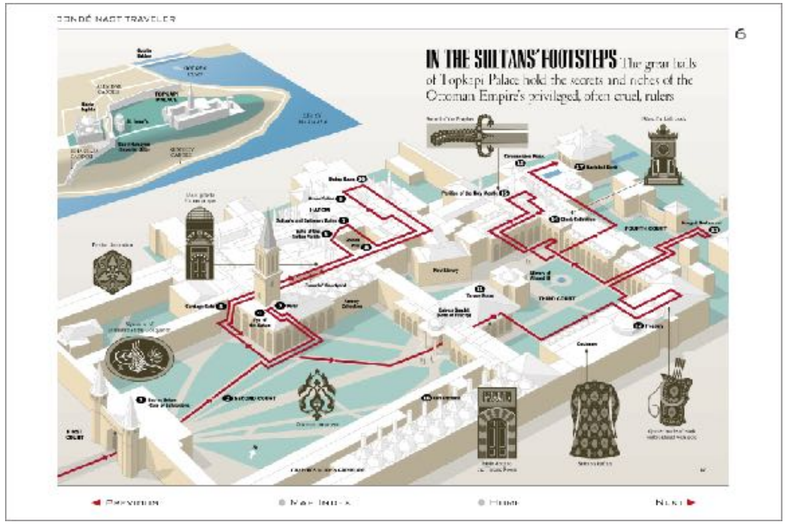



Global flow of people- Wittgenstein Centre

## Comunication

### Convey results / Explain

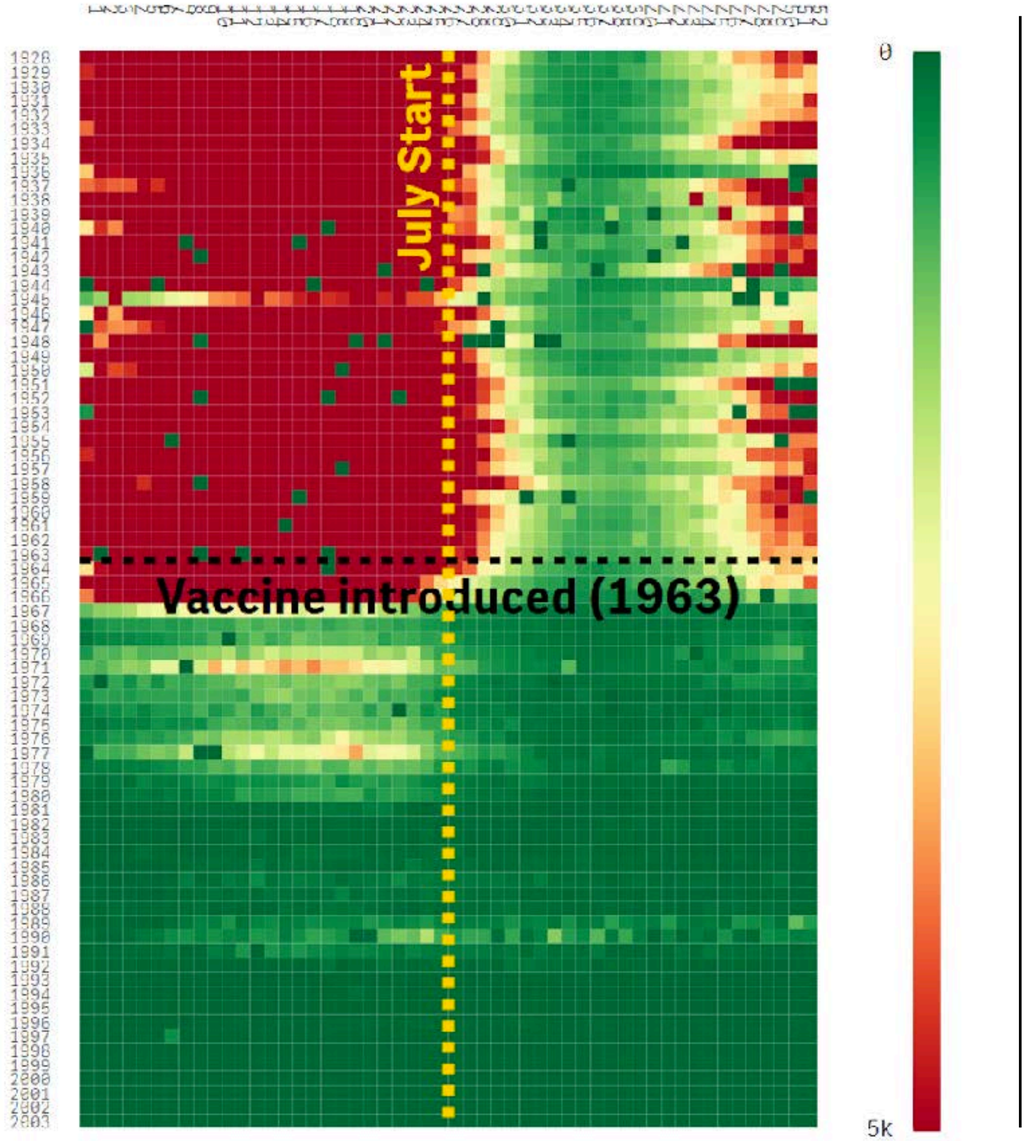
Posters
Papers
Slides

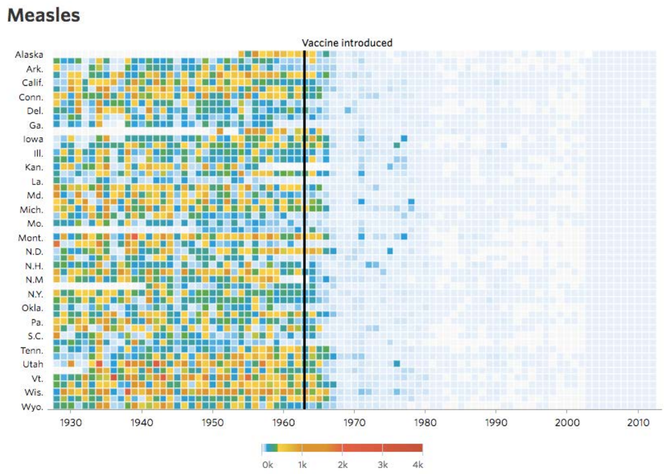


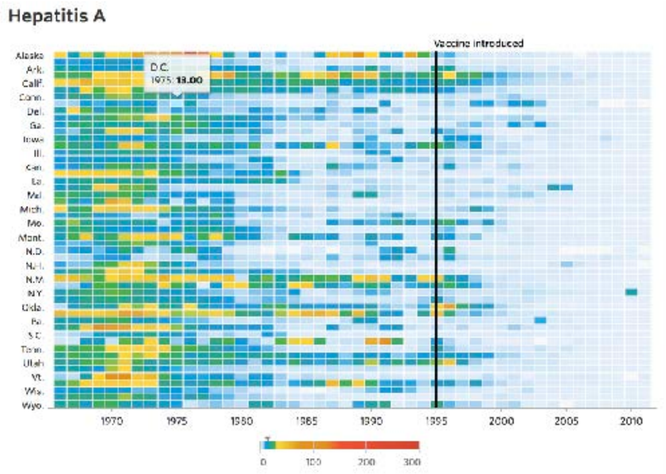

John Grimmwade

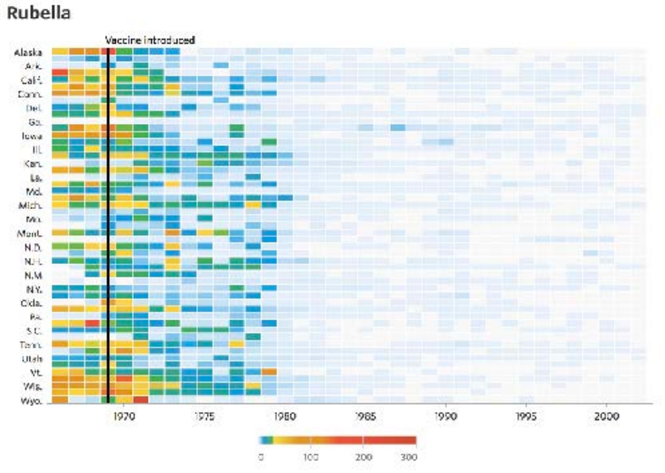
# Battling Infectious Diseases in the 20th Century: The Impact of Vaccines

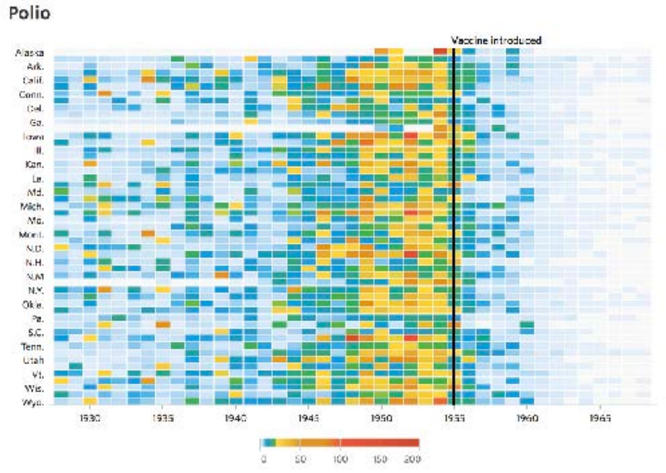
Number of Measle Incidents



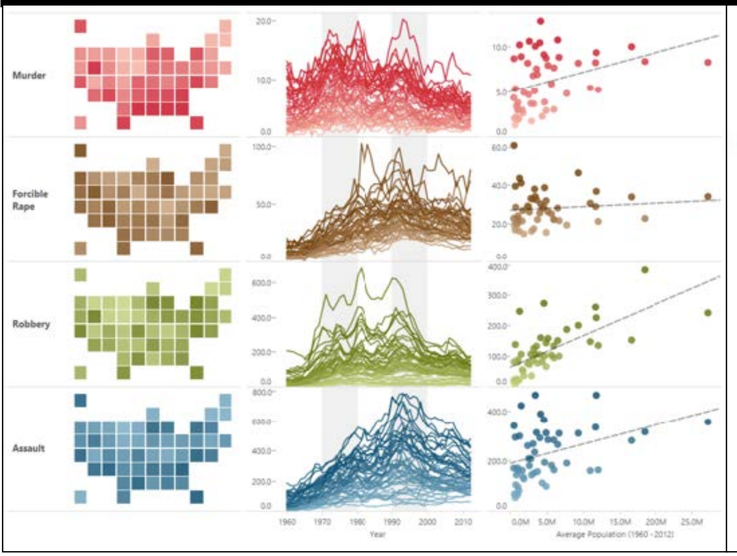




www.vizlib.com

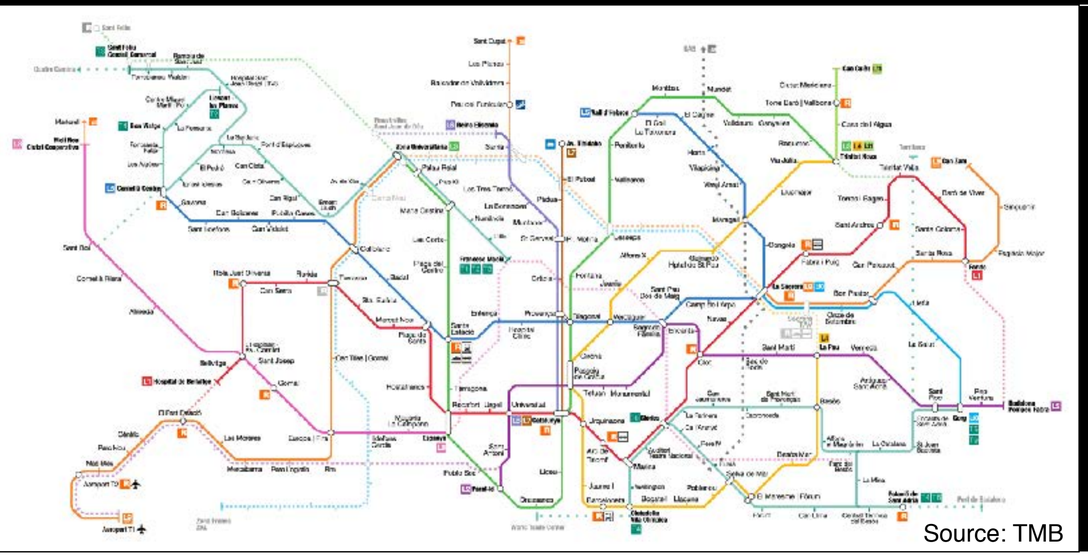
# Visualisation

## Data Viz



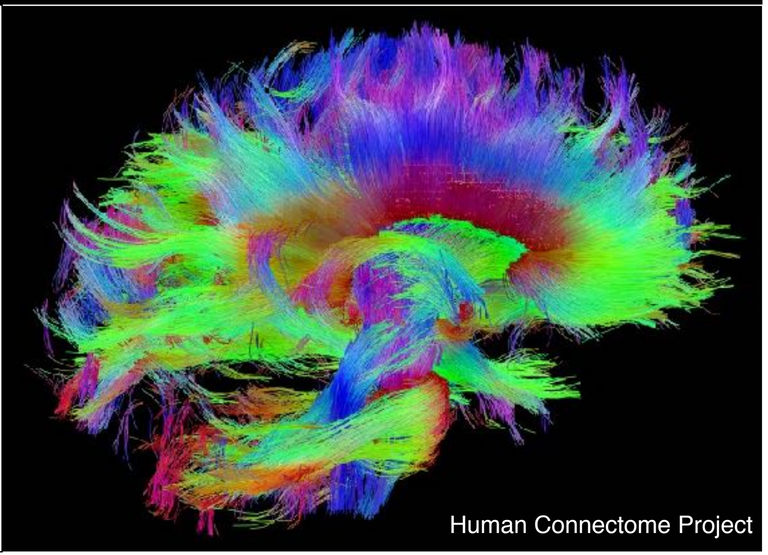

A display of any kind of data

## Information Visualisation



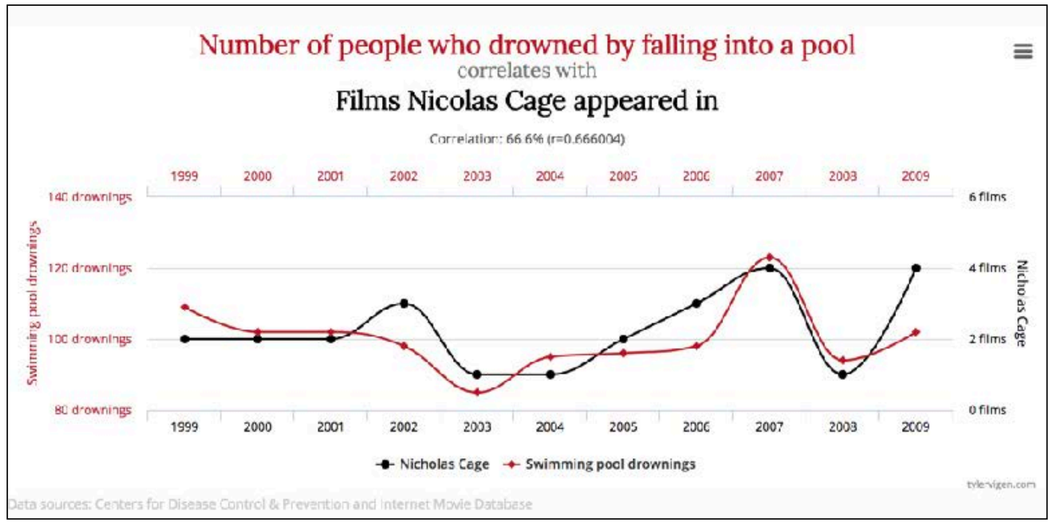

A display of other types of information

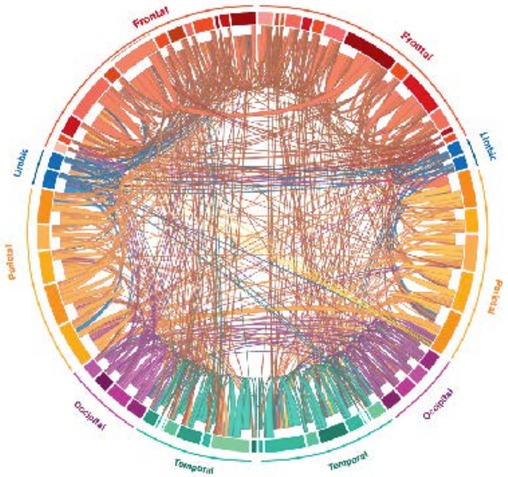
## Scientific Viz

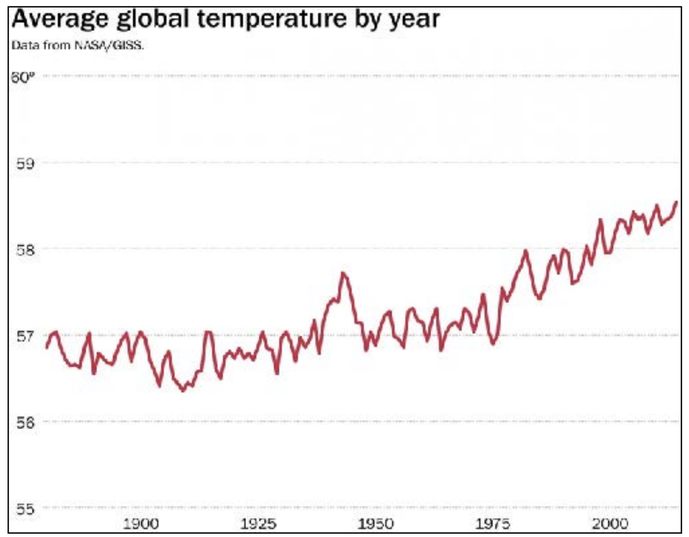


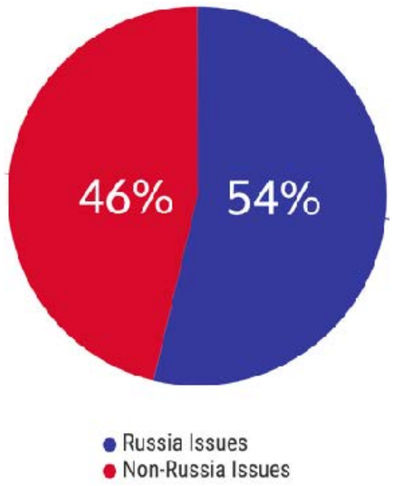

The visual presentation of scientific data
(medical, GIS, physics, etc)

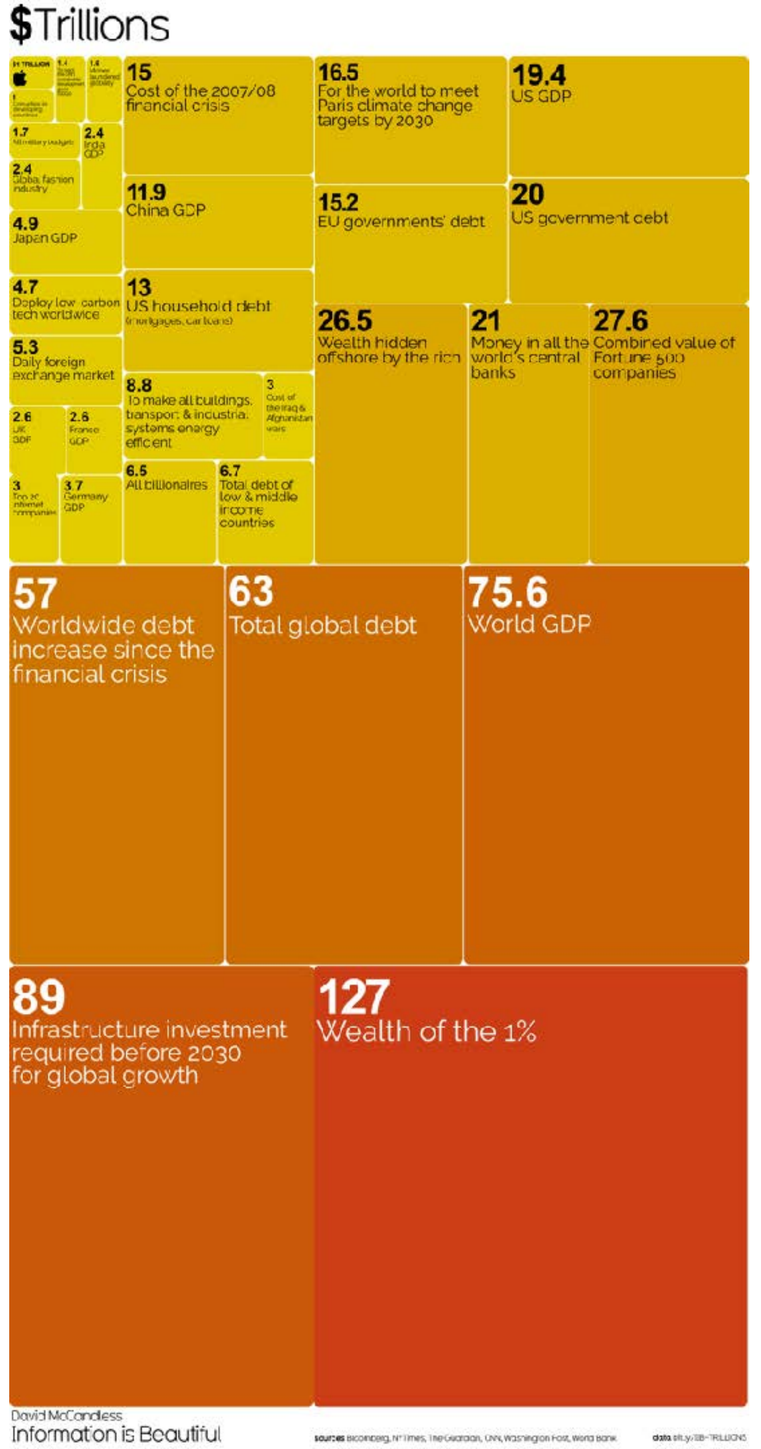
# Data Viz


Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in
Correlation: 66.6% (r=0.666004)
1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
140 drowningss
120 drowningss
100 drowningss
80 drowningss
Swimming pool drownings
6 films
4 films
2 films
0 films
Nicholas Cage
Nicholas Cage
Swimming pool drownings
tylervigen.com
Data sources: Centers for Disease Control & Prevention and Internet Movie Database

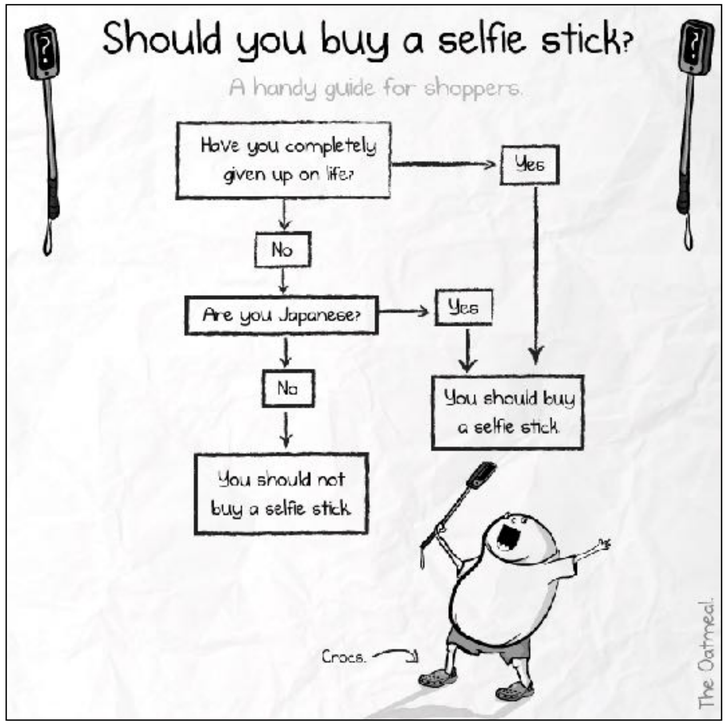

Frontal
Frontal
Limbic
Limbic
Parietal
Parietal
Occipital
Occipital
Temporal
Temporal

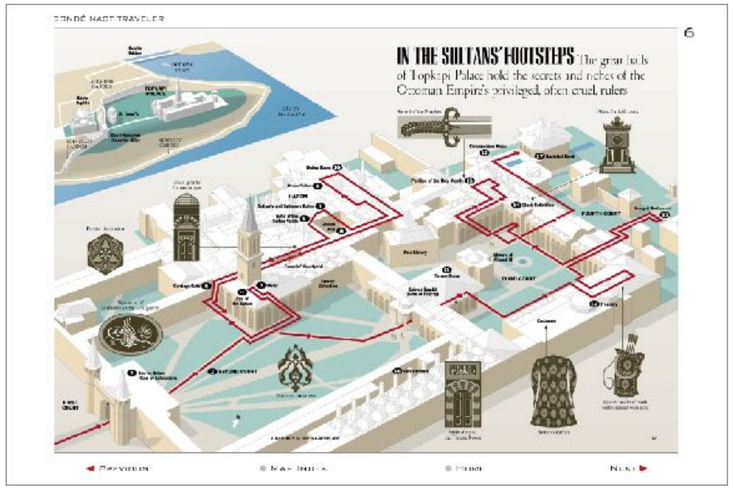

Average global temperature by year
Data from NASA/GISS.
60°
59
58
57
56
55
1900
1925
1950
1975
2000

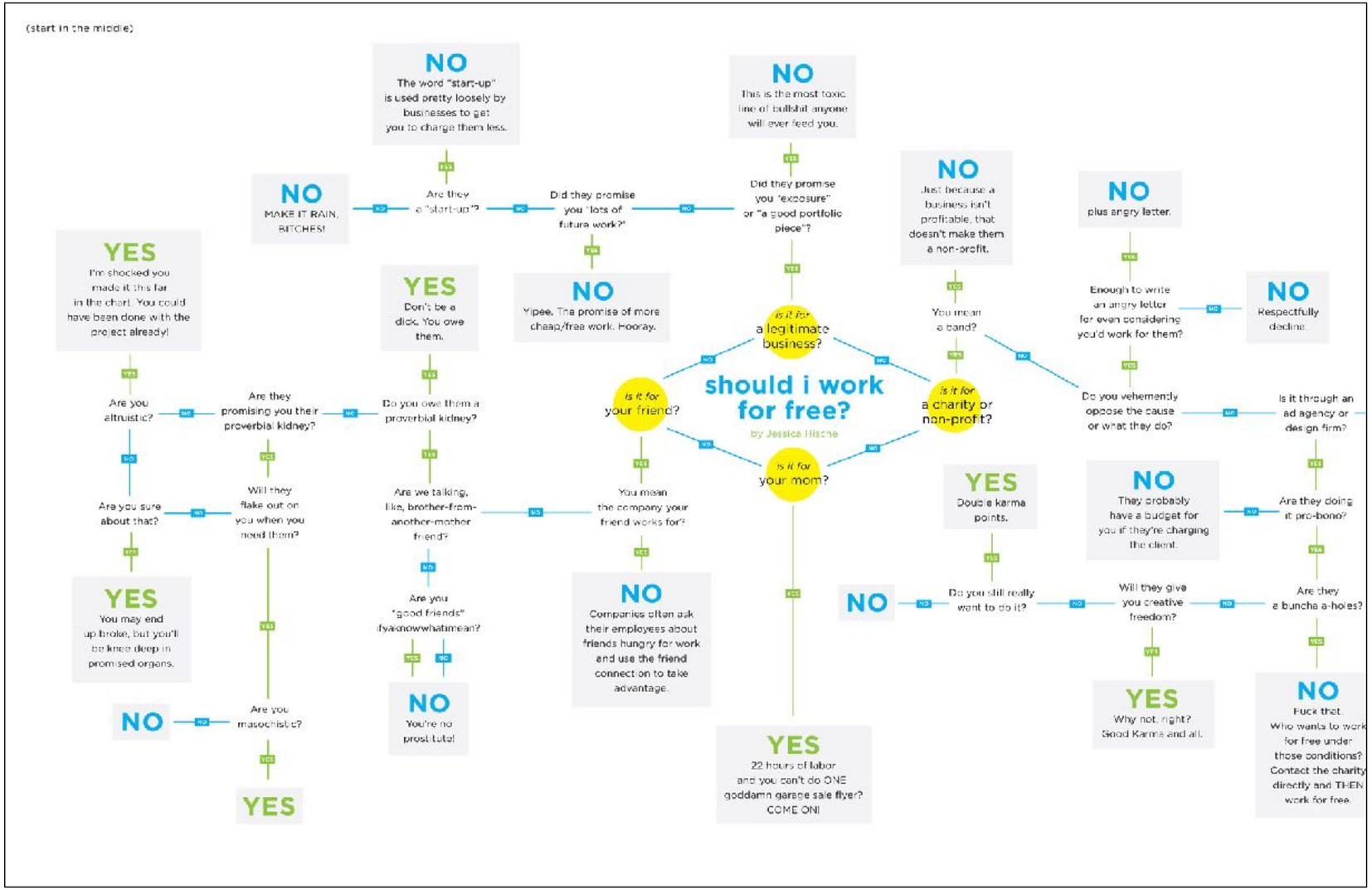

46%
54%
Russia Issues
Non-Russia Issues


$Trillions
15
Cost of the 2007/08 financial crisis
16.5
For the world to meet Paris climate change targets by 2030
19.4
US GDP
11.9
China GDP
15.2
EU governments' debt
20
US government debt
4.9
Japan GDP
13
US household debt
4.7
Deploy low carbon tech worldwide
26.5
Wealth hidden offshore by the rich
21
Money in all the world's central banks
27.6
Combined value of Fortune 500 companies
5.3
Daily foreign exchange market
8.8
To make all buildings, transport & industria systems energy efficient
6.5
All billionaires
6.7
Total debt of low & middle income countries
57
Worldwide debt increase since the financial crisis
63
Total global debt
75.6
World GDP
89
Infrastructure investment required before 2030 for global growth
127
Wealth of the 1%
David McCandless
Information is Beautiful

# Info Viz

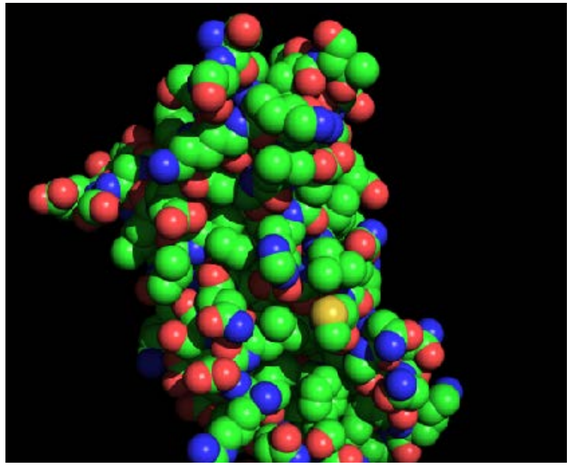
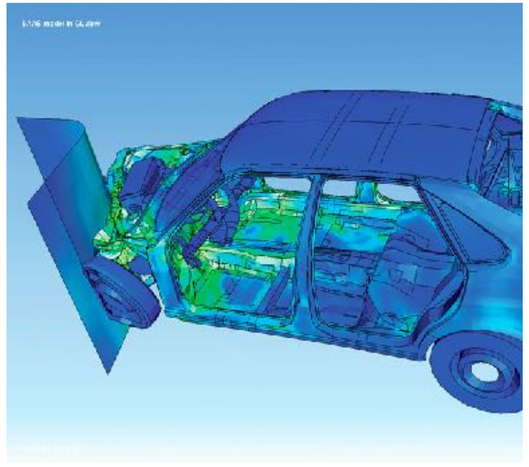
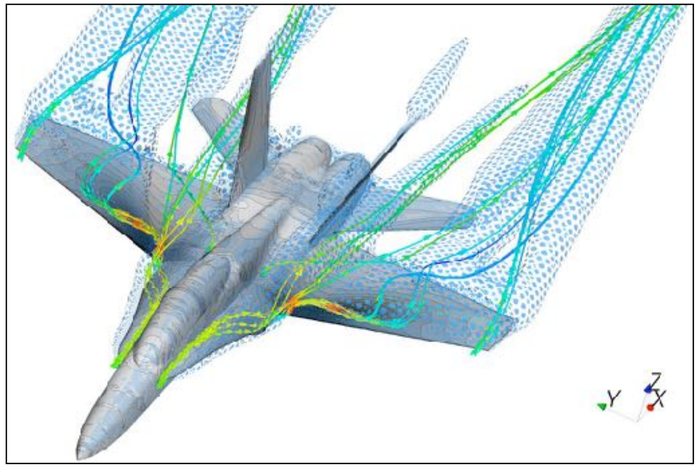
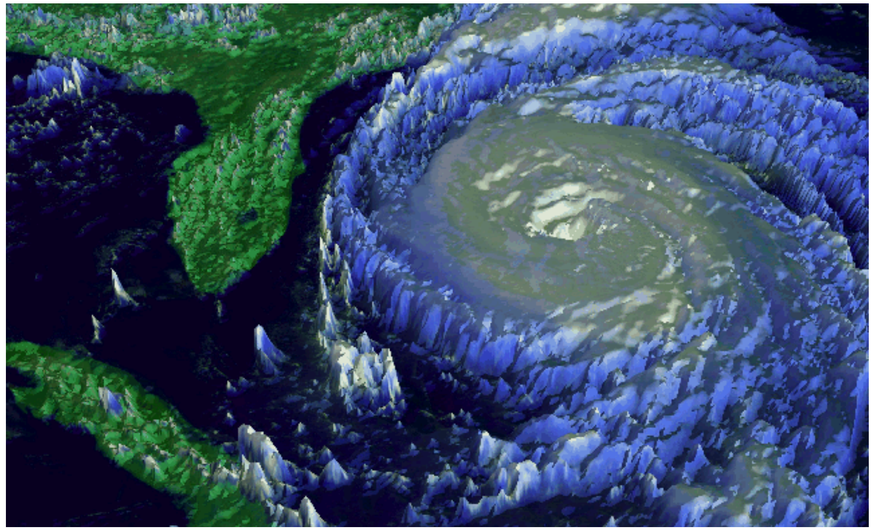
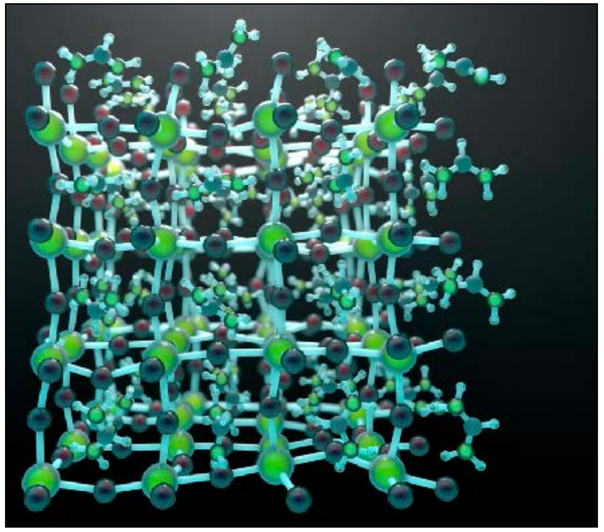
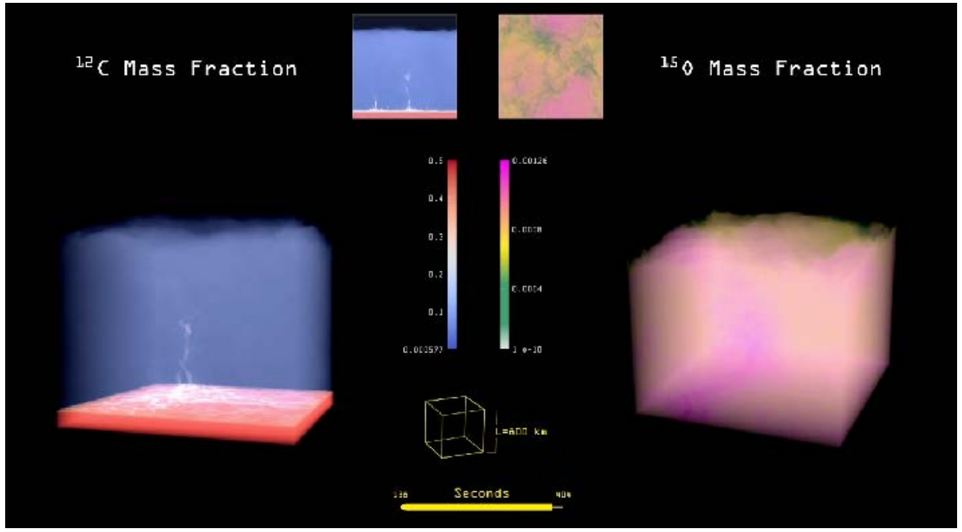

Should you buy a selfie stick?
A handy guide for shoppers.
Have you completely given up on life?
Yes
No
Are you Japanese?
Yes
No
You should buy a selfie stick.
You should not buy a selfie stick.
Crocs.
The Oatmeal


CONDÉ NAST TRAVELER
6
IN THE SULTANS' FOOTSTEPS The great halls of Topkapi Palace hold the secrets and riches of the Ottoman Empire's privileged, often cruel, rulers


(start in the middle)
should i work for free?
by Jessica Hische
is it for your friend?
is it for a charity or non-profit?
is it for a legitimate business?
is it for your mom?
NO
The word "start-up" is used pretty loosely by businesses to get you to charge them less.
NO
MAKE IT RAIN, BITCHES!
Are they a "start-up"?
Did they promise you "lots of future work?"
NO
Yipee. The promise of more cheap/free work. Hooray.
NO
This is the most toxic line of bullshit anyone will ever feed you.
Did they promise you "exposure" or "a good portfolio piece"?
YES
I'm shocked you made it this far in the chart. You could have been done with the project already!
Are you altruistic?
Are you sure about that?
YES
You may end up broke, but you'll be knee deep in promised organs.
Are they promising you their proverbial kidney?
Will they flake out on you when you need them?
NO
Are you masochistic?
YES
YES
Don't be a dick. You owe them.
Do you owe them a proverbial kidney?
Are we talking, like, brother-from-another-mother friend?
Are you "good friends" (fyaknowwhatimean?
NO
You're no prostitute!
You mean the company your friend works for?
NO
Companies often ask their employees about friends hungry for work and use the friend connection to take advantage.
YES
22 hours of labor and you can't do ONE goddamn garage sale flyer? COME ON!
NO
Just because a business isn't profitable, that doesn't make them a non-profit.
You mean a band?
NO
plus angry letter.
Enough to write an angry letter for even considering you'd work for them?
NO
Respectfully decline.
Do you vehemently oppose the cause or what they do?
Is it through an ad agency or design firm?
YES
Double karma points.
NO
Do you still really want to do it?
Will they give you creative freedom?
YES
Why not, right? Good Karma and all.
NO
They probably have a budget for you if they're charging the client.
Are they doing it pro-bono?
Are they a buncha a-holes?
NO
Fuck that. Who wants to work for free under those conditions? Contact the charity directly and THEN work for free.

# Scientific Viz

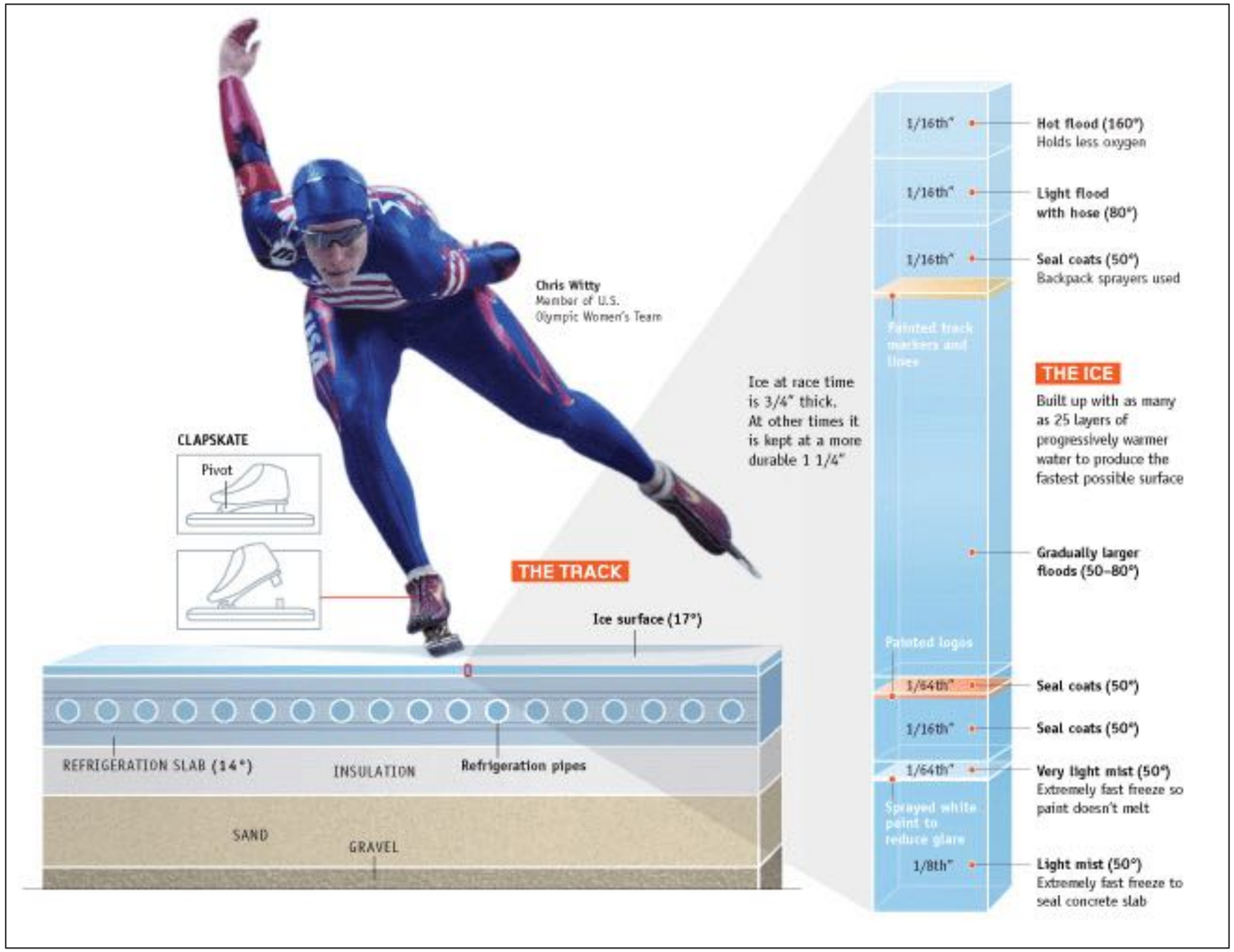
- Medical Imaging
- Math
- Simulations of physical phonemena
- Geographical Information Systems

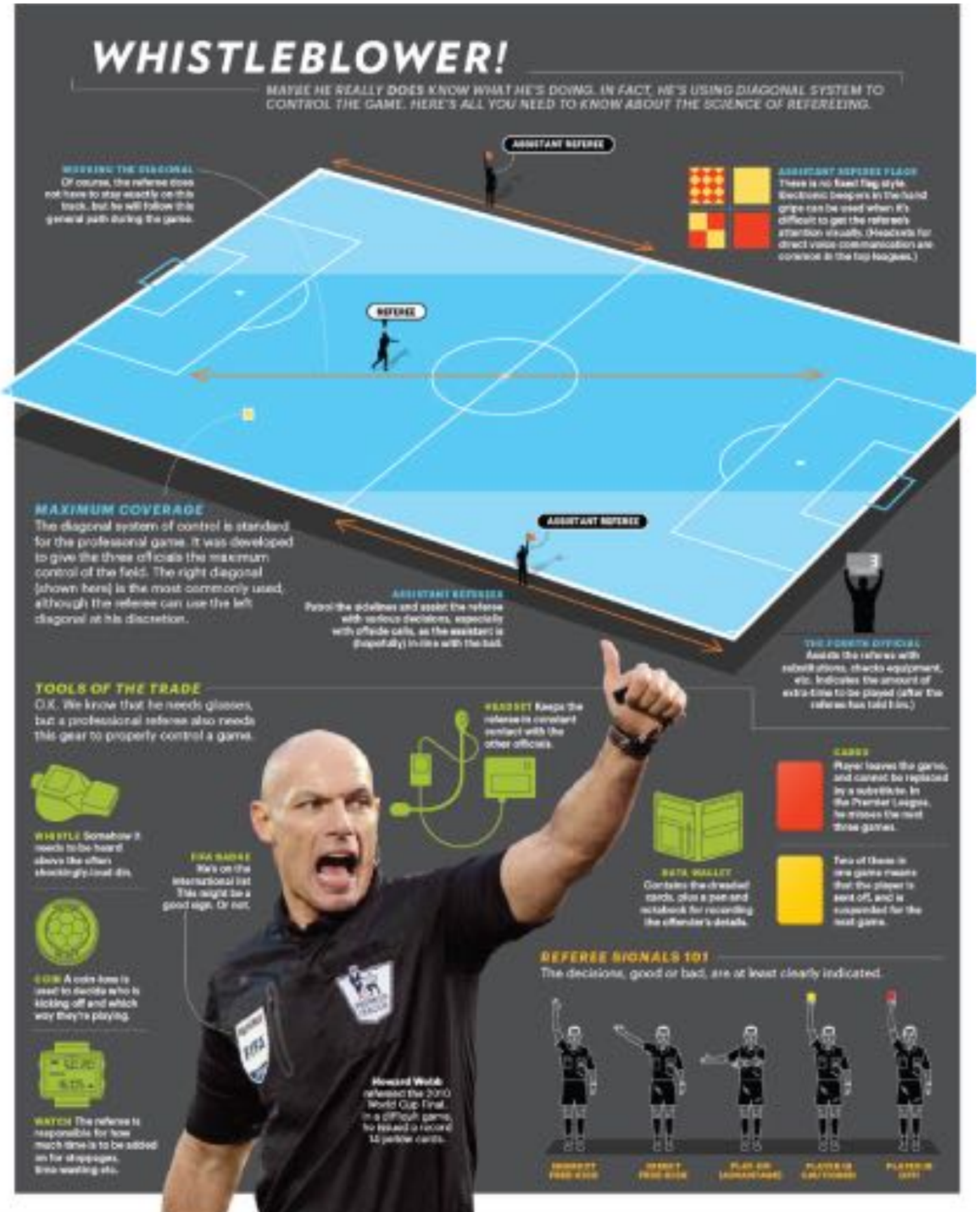
# Data everywhere, visualisations everywhere

# Infographics

Multi-section visual representation of information intended to communicate one or more specific messages

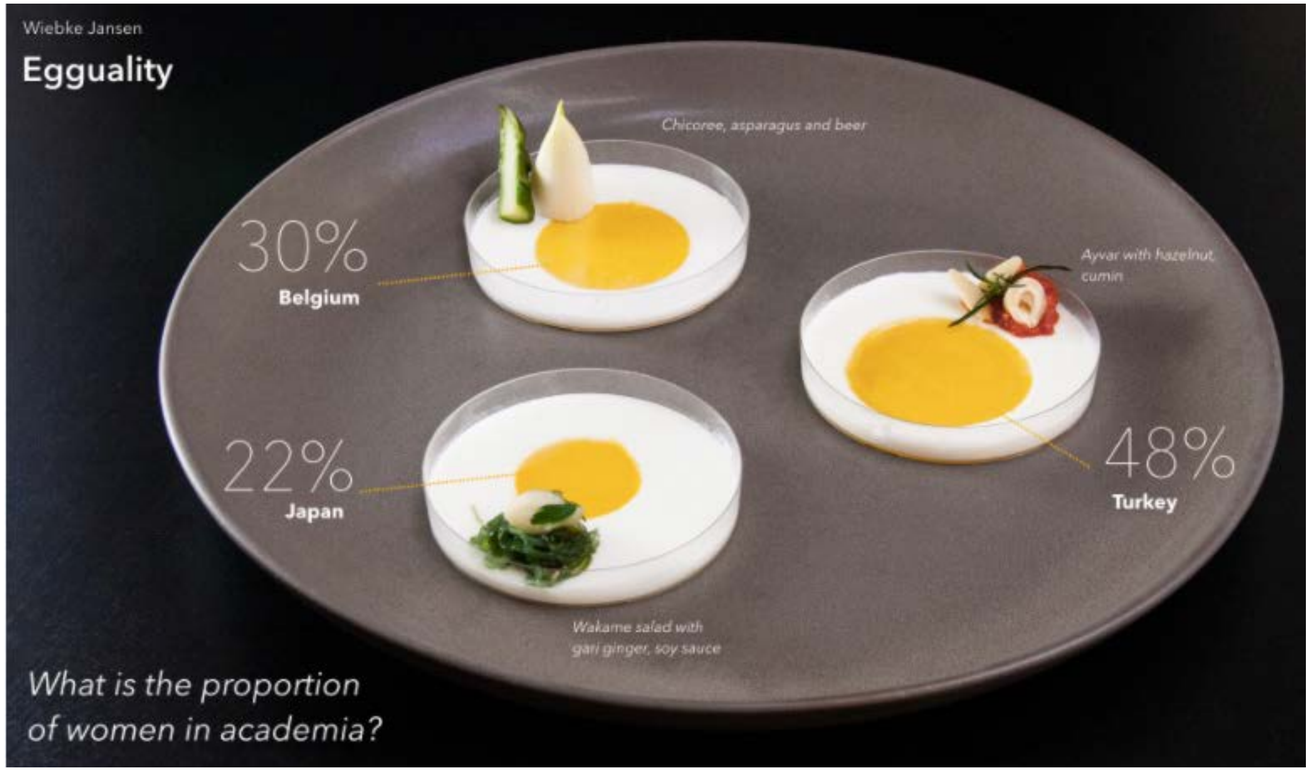



John Grimwade



# Data physicalisation

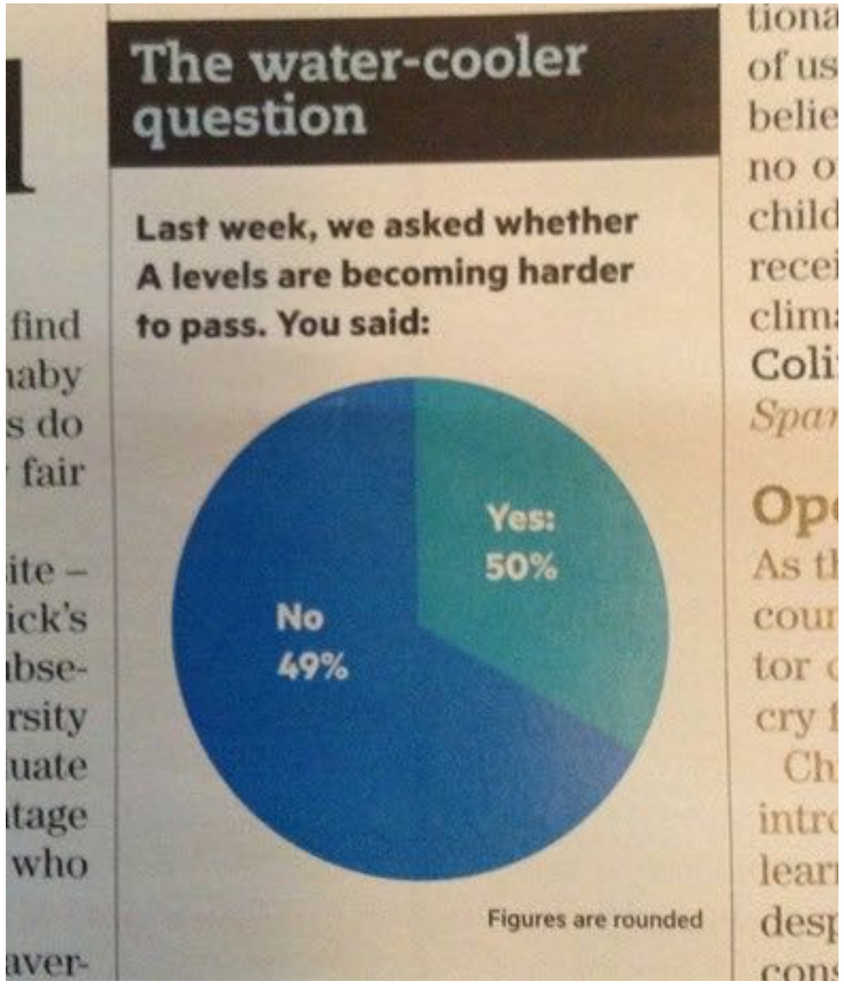
# Data food

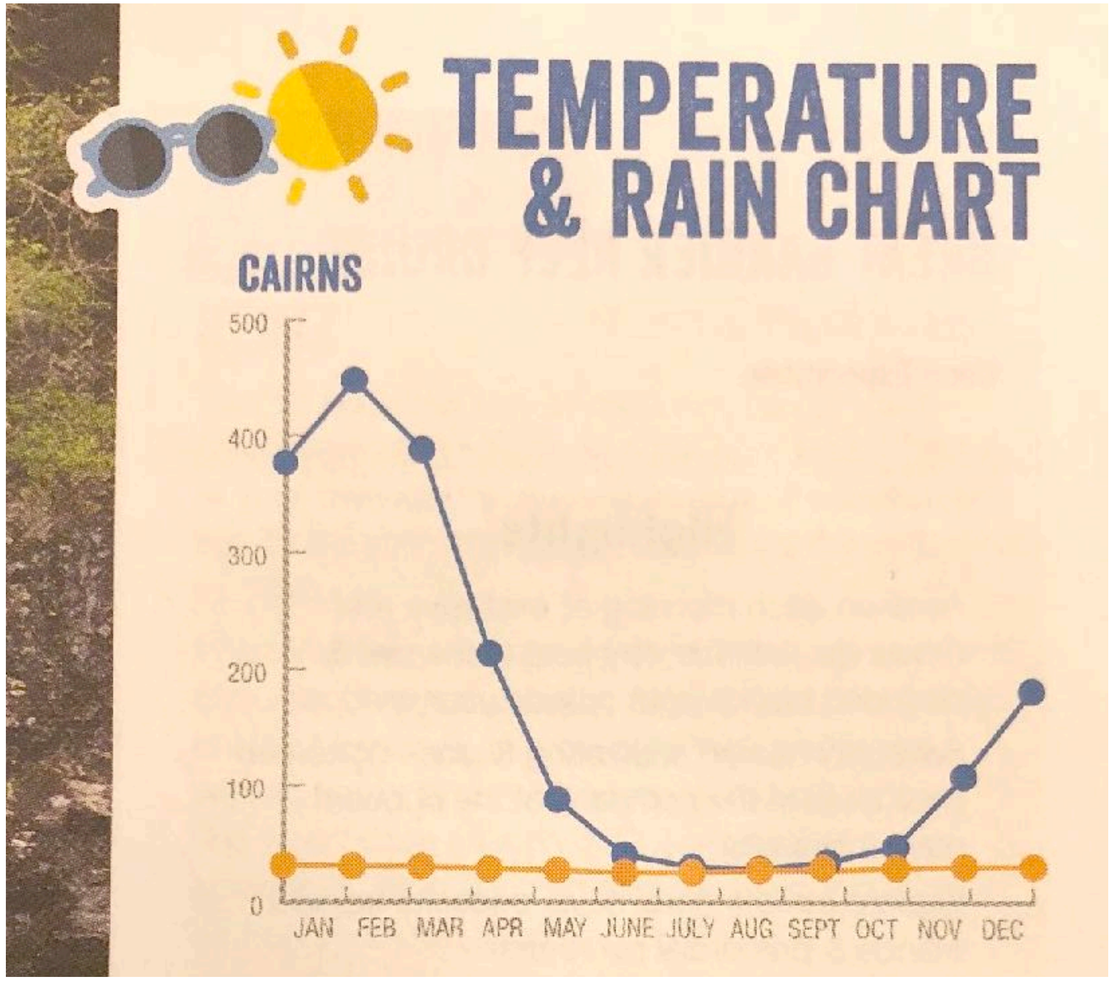


# Who cares?

Defaults don’t work in all cases, and software don’t fit all data

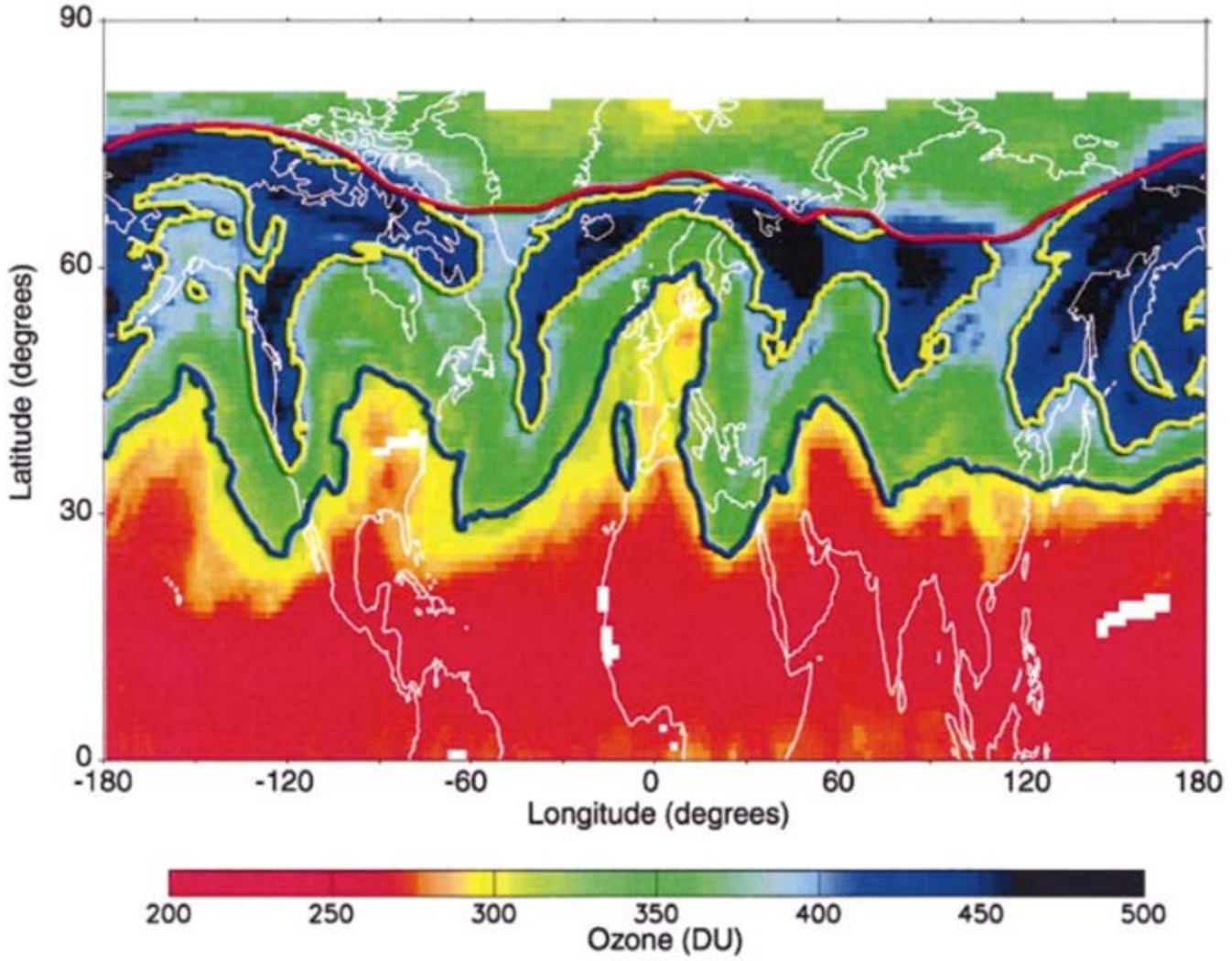
A wrong choice of chart or design decisions can lead to mistakes



Even to distort the message

# Even to distort the message



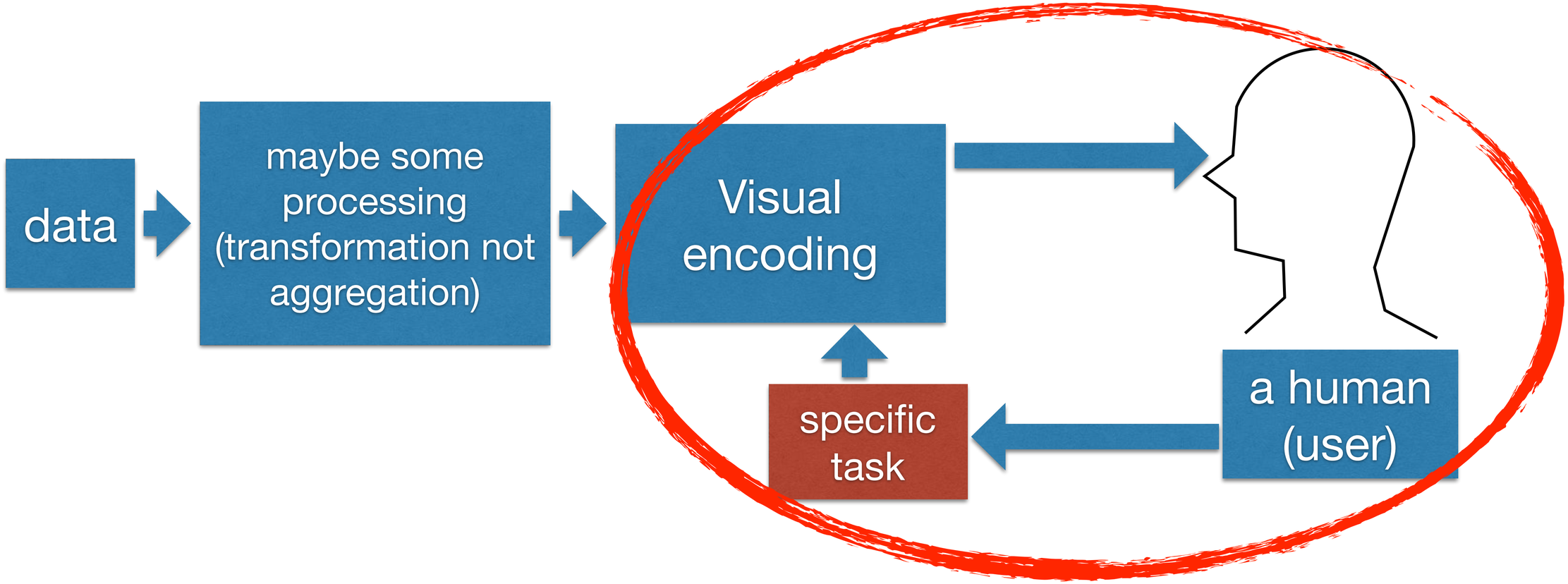

Hudson, R. D., Andrade, M. F., Follette, M. B., and Frolov, A. D.:
The total ozone field separated into meteorological regimes – Part II: Northern Hemisphere mid-latitude total ozone trends,
Atmos. Chem. Phys., 6, 5183-5191, https://doi.org/10.5194/acp-6-5183-2006, 2006.

# What is visualisation?

Cognitive tool to enable analysis, exploration, and discovery

Focused on user/task

Mixture of other disciplines (cartography, statistics, graphic design, neuroscience, computer science)




(T.Munzner)

# Summary

Data, charts and visual encoding

Visual Honesty

Graphic design

Storytelling with data

# Data, charts and visual encoding

Data types by structure

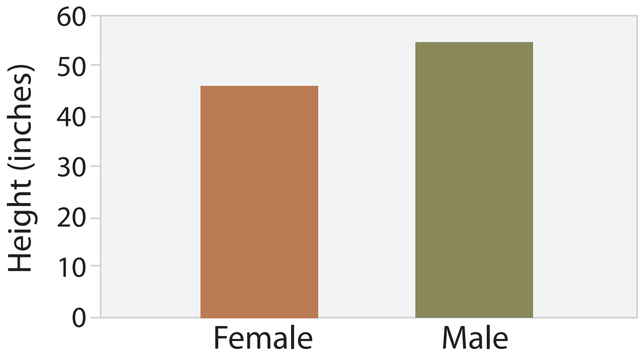
Chart types by function

Encoding: Visually represent data

# Data types

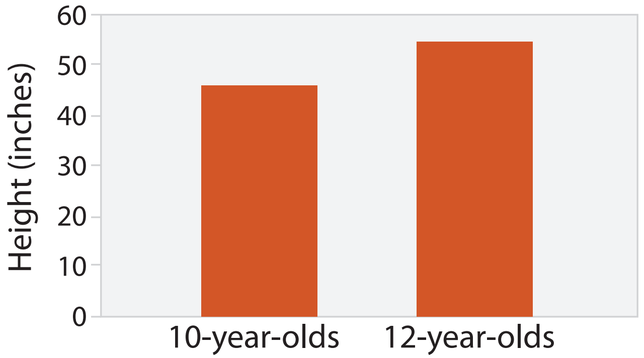
## Categorical

No quantitative relation
No inherent order



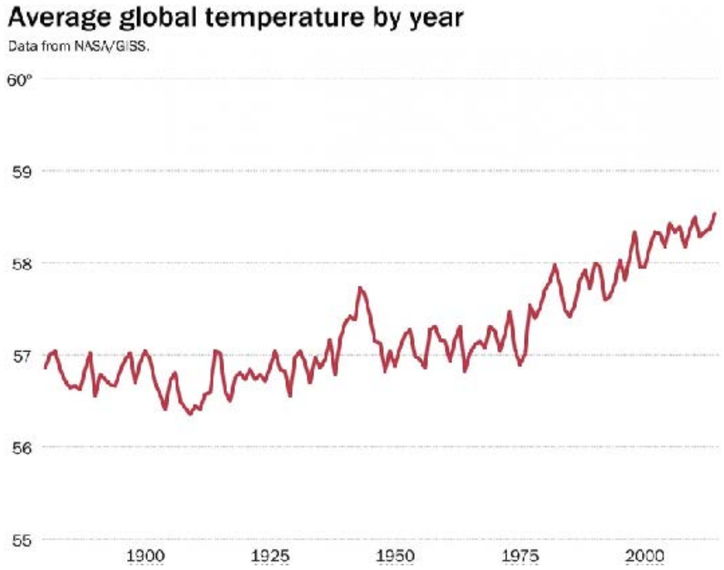

## Ordinal

Can be ordered
Degree of difference not measurable



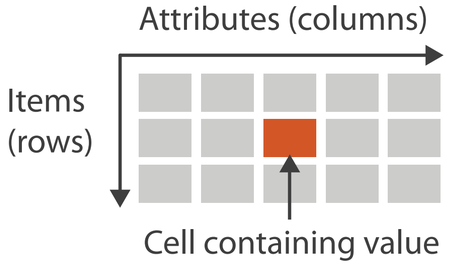

## Quantitative

Can be measured
and manipulated numerically

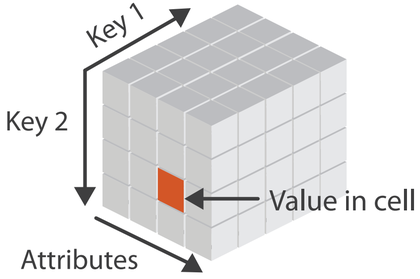


# Data types by structure

## Tables

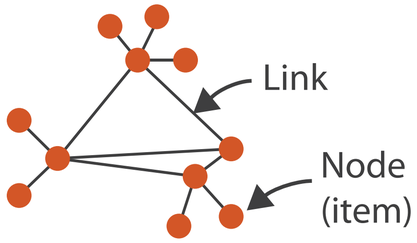



### Multidimensional

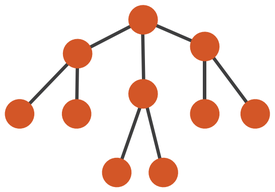



## Relational

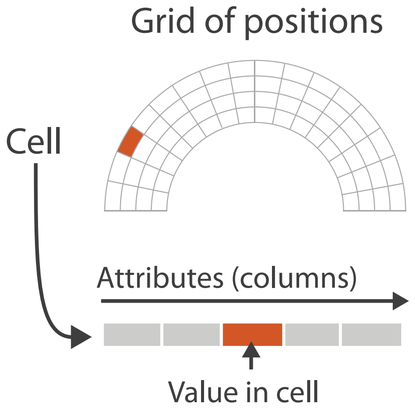
### Redes




### Hierarchy/Trees

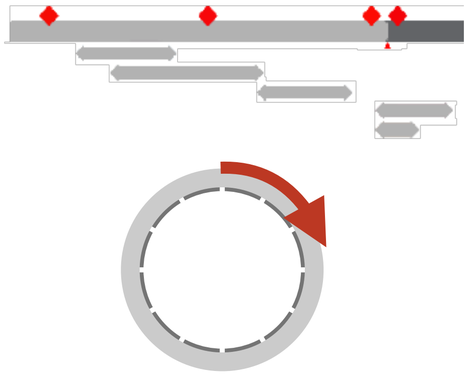


## Spatial

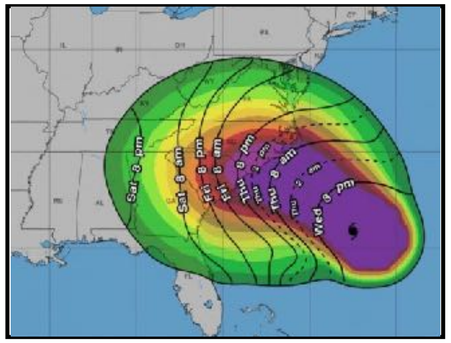



### Geometry

## Temporal



## Spatio-Temporal



## Textual

# Visual Encoding

Represent data through visual channels

Draw 37 and 73

# Visual Channels

Position on common scale

Position on unaligned scale

Length (1D size)

Tilt/angle

Area (2D size)

Depth (3D position)

Color luminance

Color saturation

Curvature

Volume (3D size)

Spatial region

Color hue

Motion

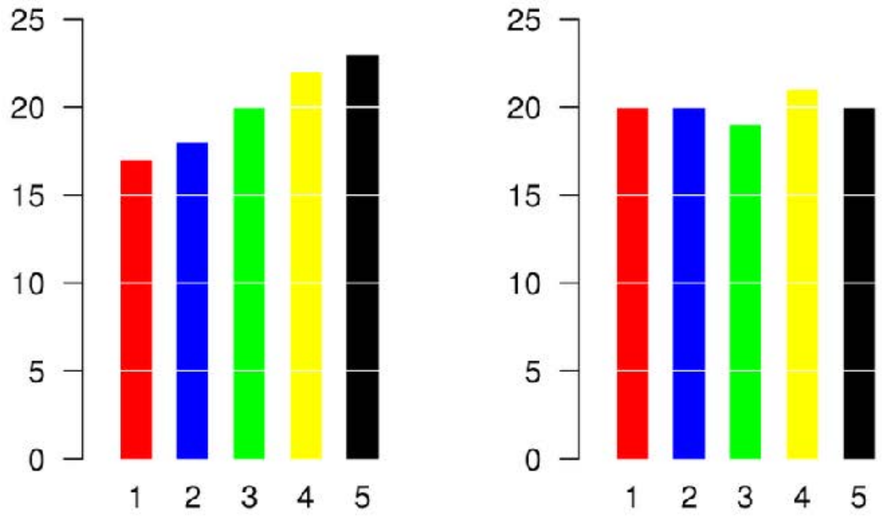
Shape

from The Truthful Art. Cairo, A.

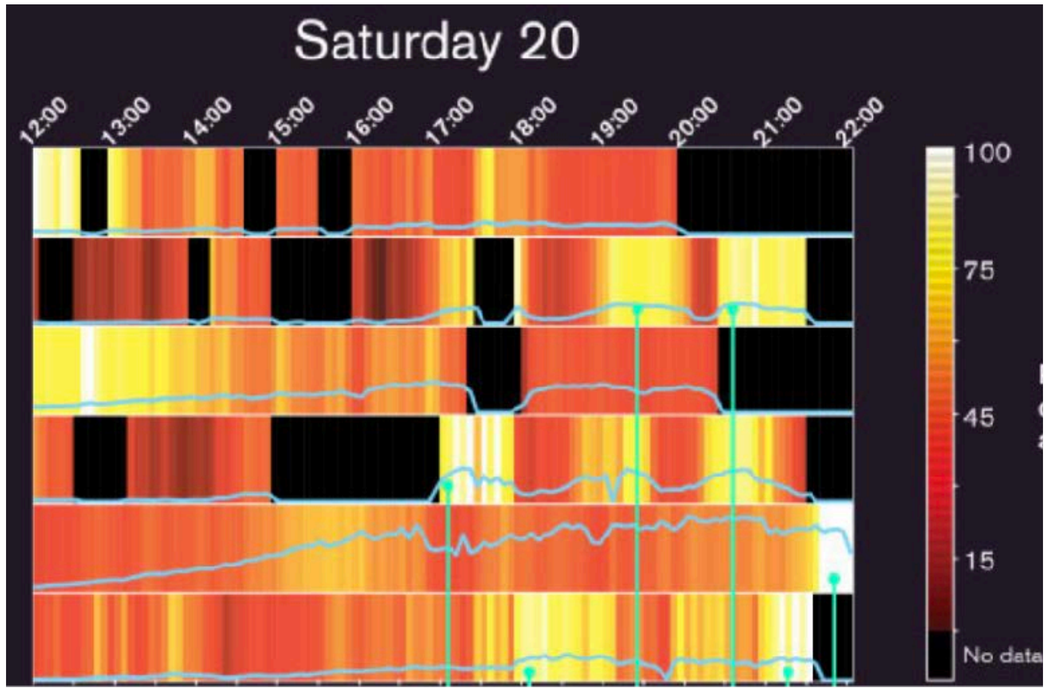
# Visual Channels

The visual channel that we choose must permit the desired analysis and comparisons

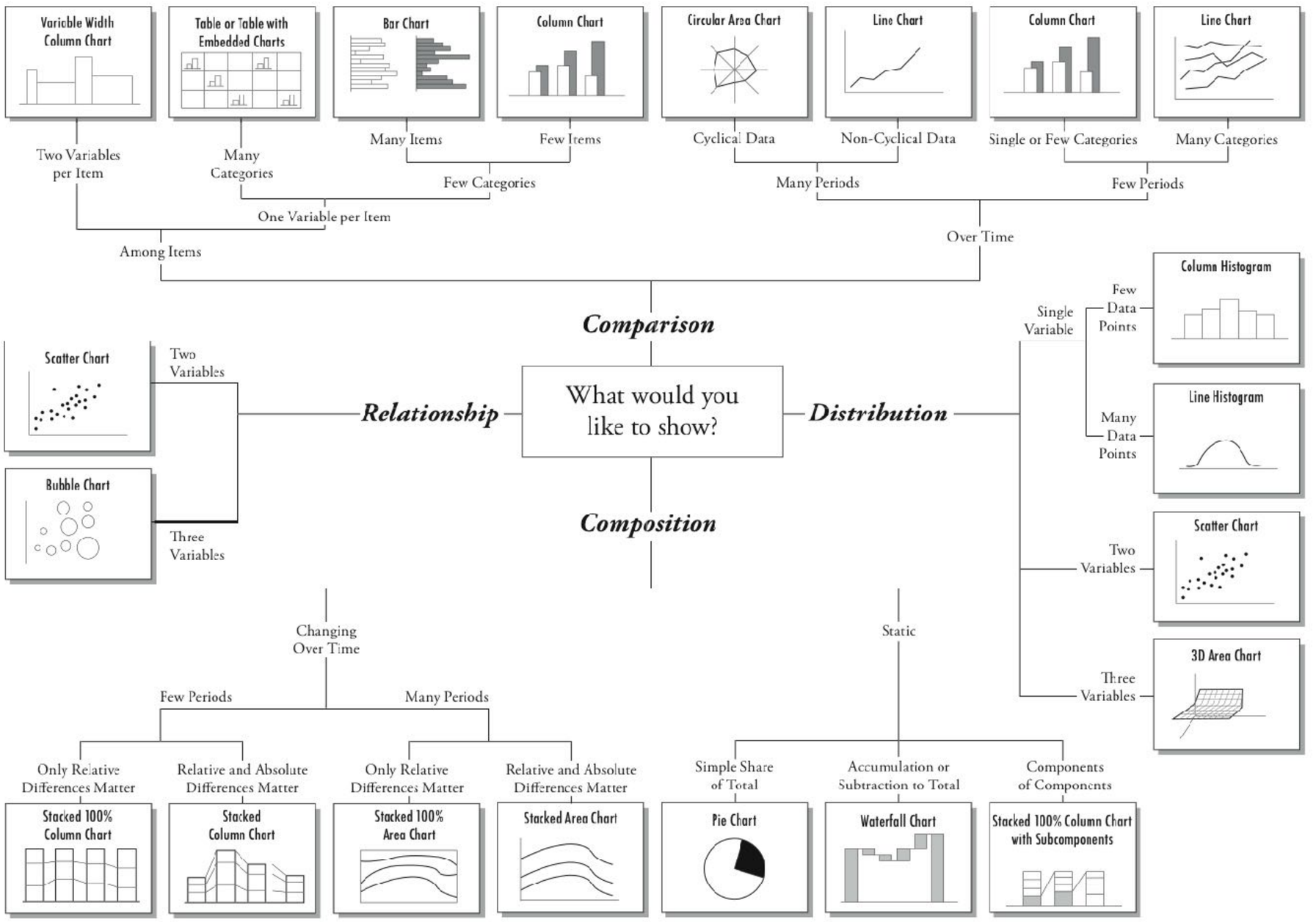
Compare categorical values
Unrelated categories = different colours




Progressive variation in data = in color



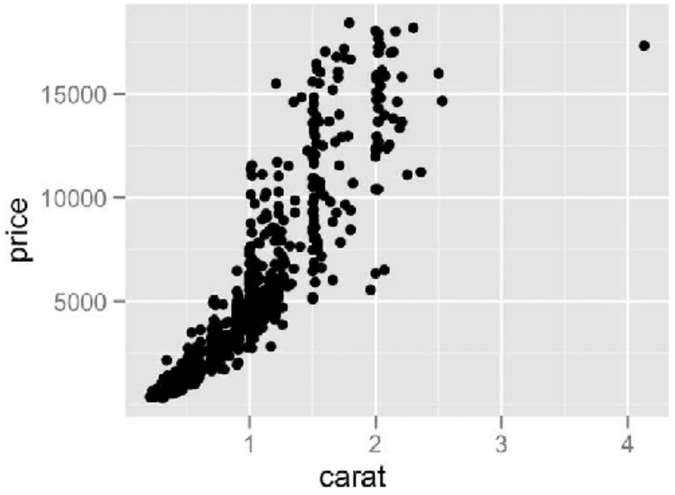
# Charts

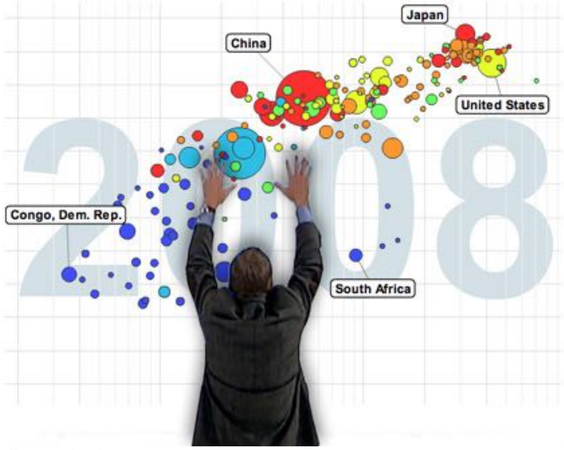



www.ExtremePresentation.com
© 2009 A. Abela — a.v.abela@gmail.com

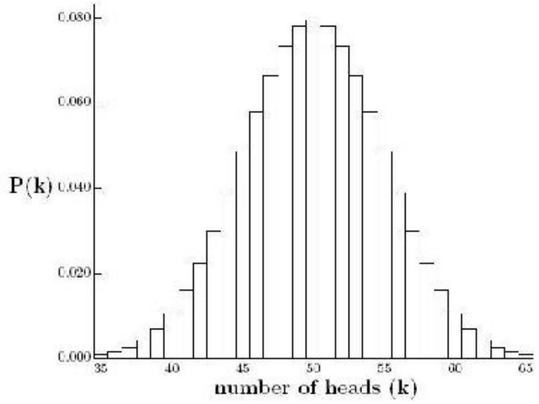
# Types of charts

## Scatter plot - Bubble chart - histogram



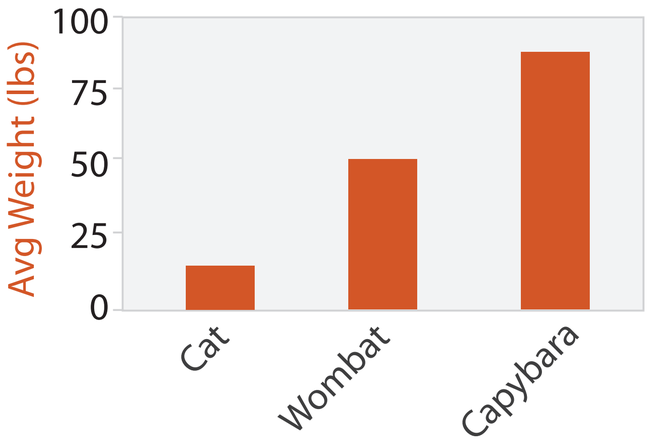



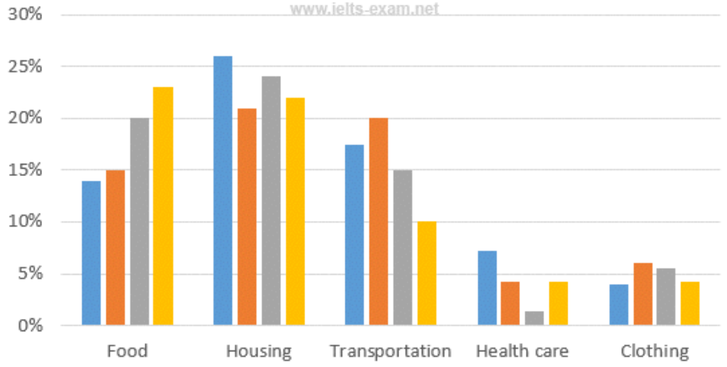

Gapminder

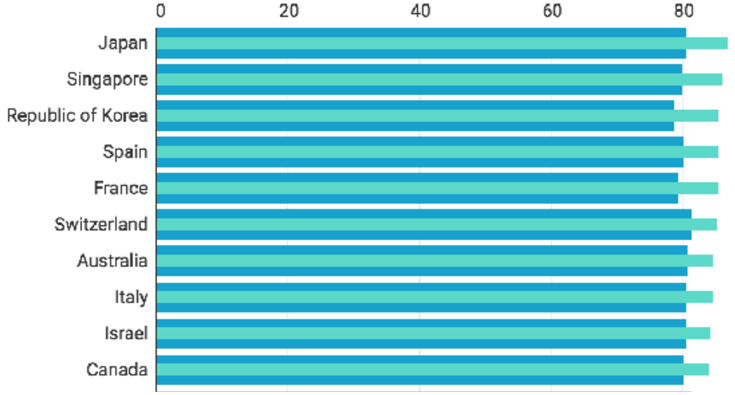



David Terr

## Bars - Grouped bars - Bullet plot

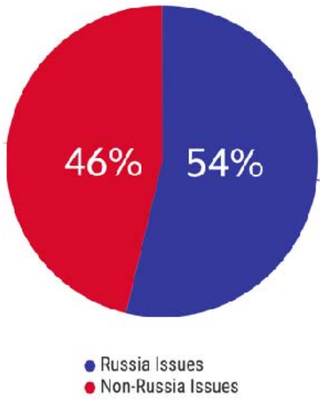


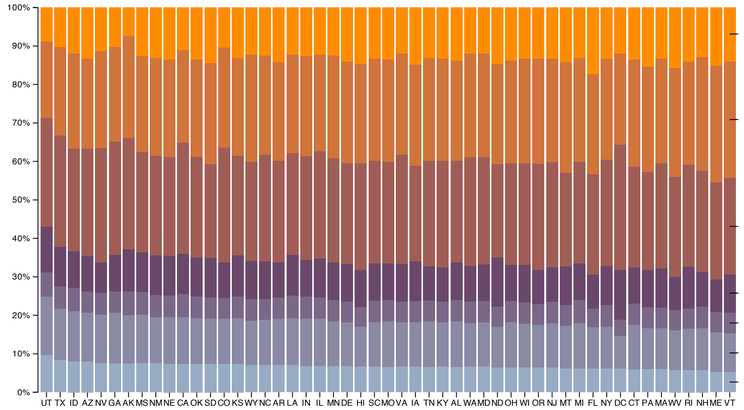



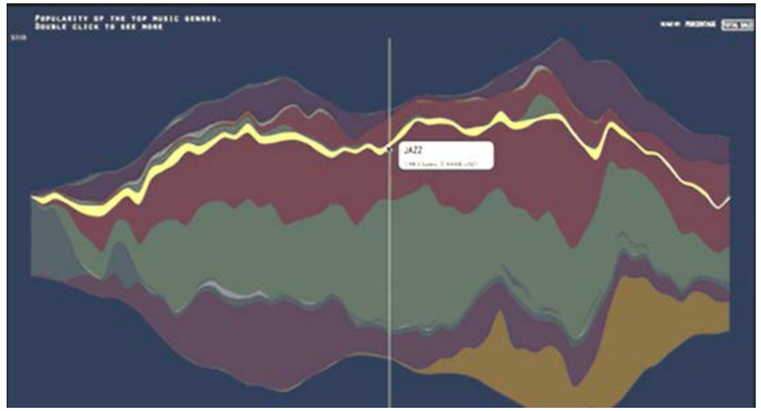



# Types of charts

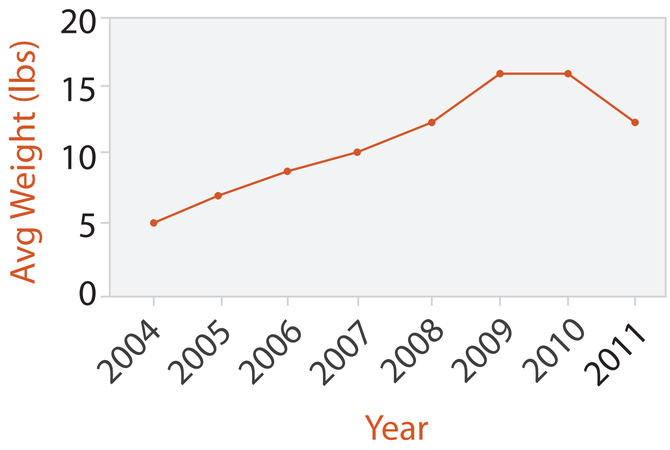
## Pie chart - stacked bars - Stream graph

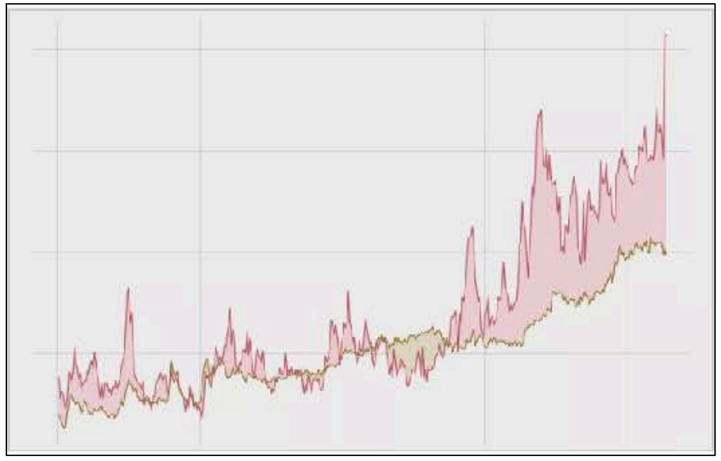


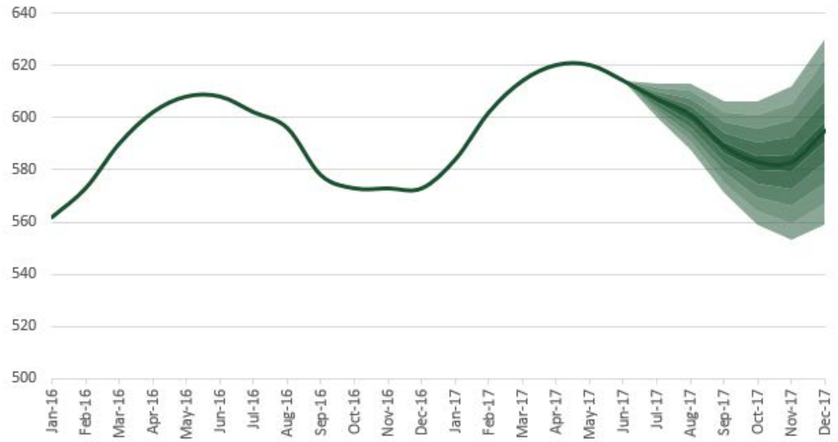



## Line chart - Difference chart - Fan chart



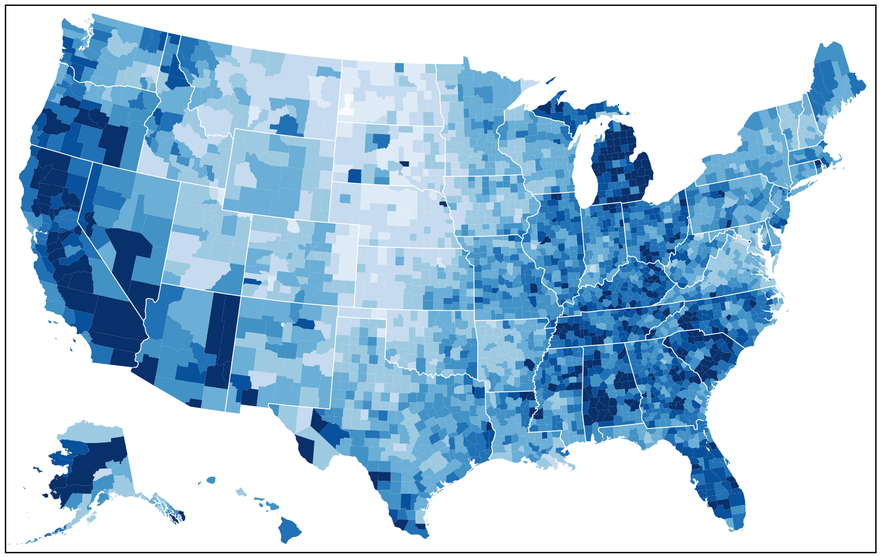
https://flowingdata.com/charttype/difference-chart/



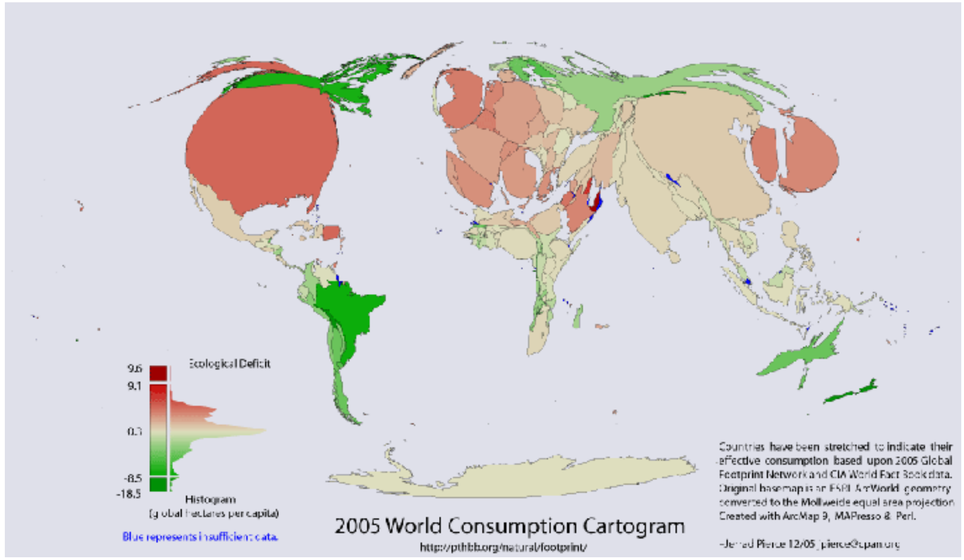

exceloffthegrid.com

# Types of charts

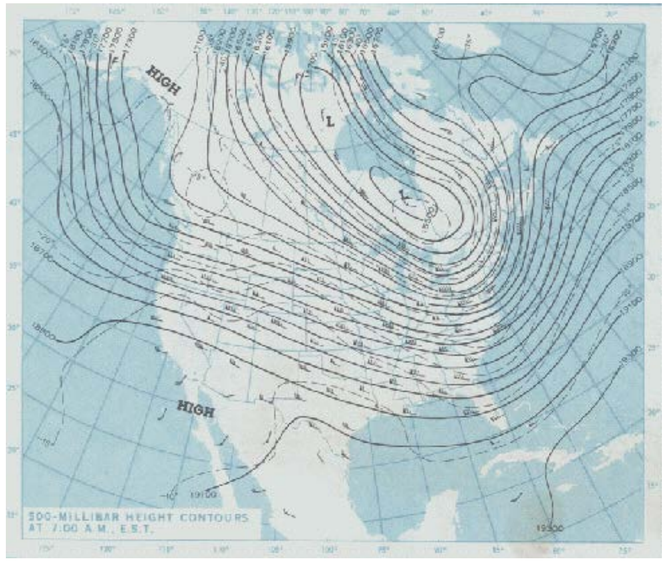
Choropleths- cartograms - contour maps



http://bl.ocks.org/mbostock/4060606



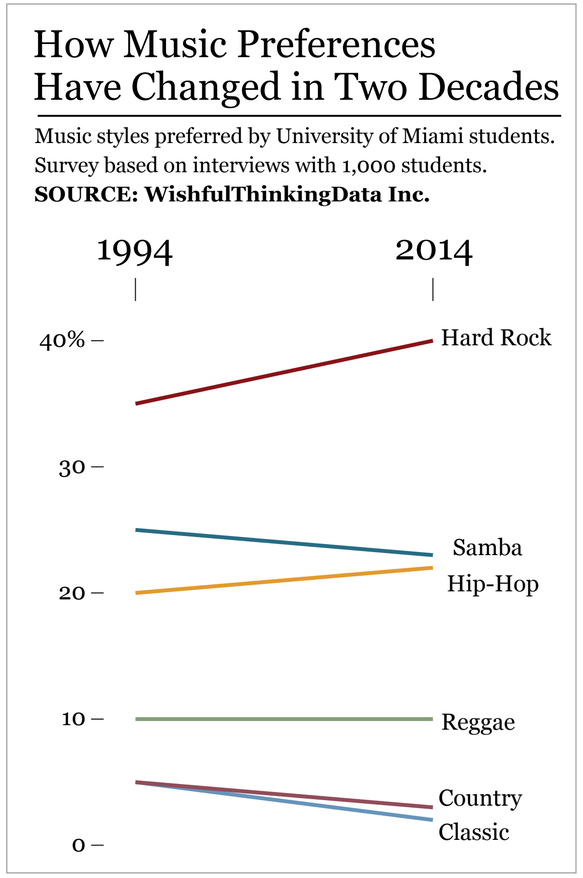

from: www.pthbb.org/natural/footprint/

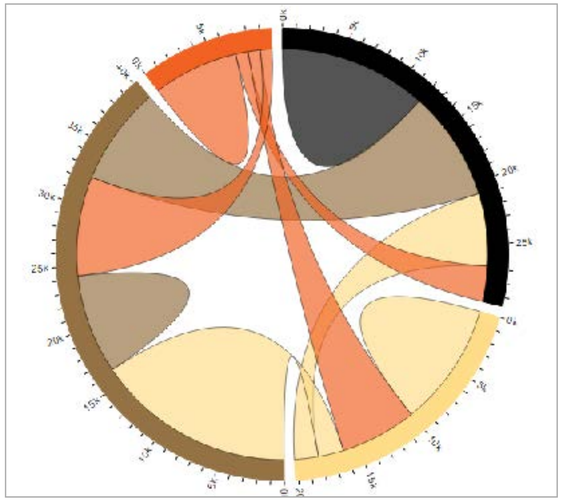



NOAA Central Library, Silver Spring, Maryland

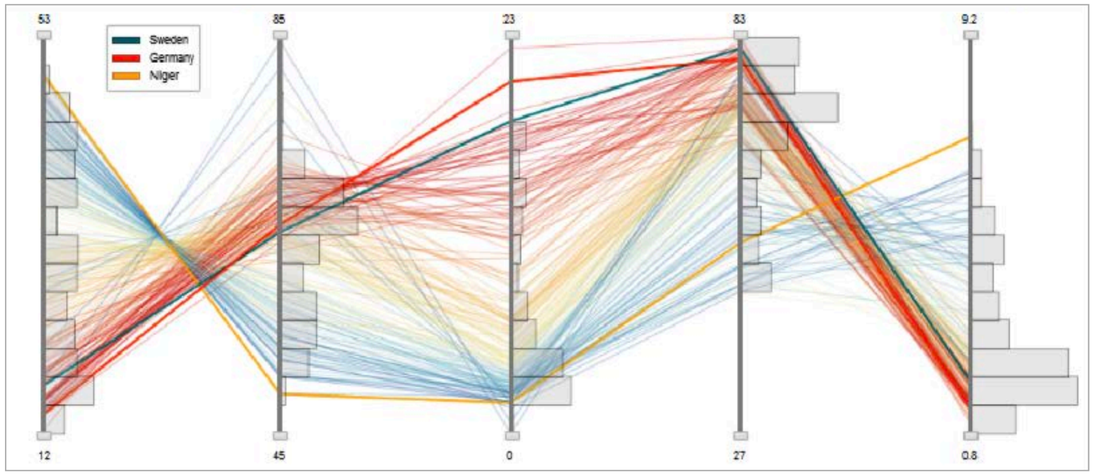
# Types of charts

## Slope graph - Chord diagram - Parallel coordinates - treemap



d3.js




from: ncva.itn.liu.se/education-geovisual-analytics

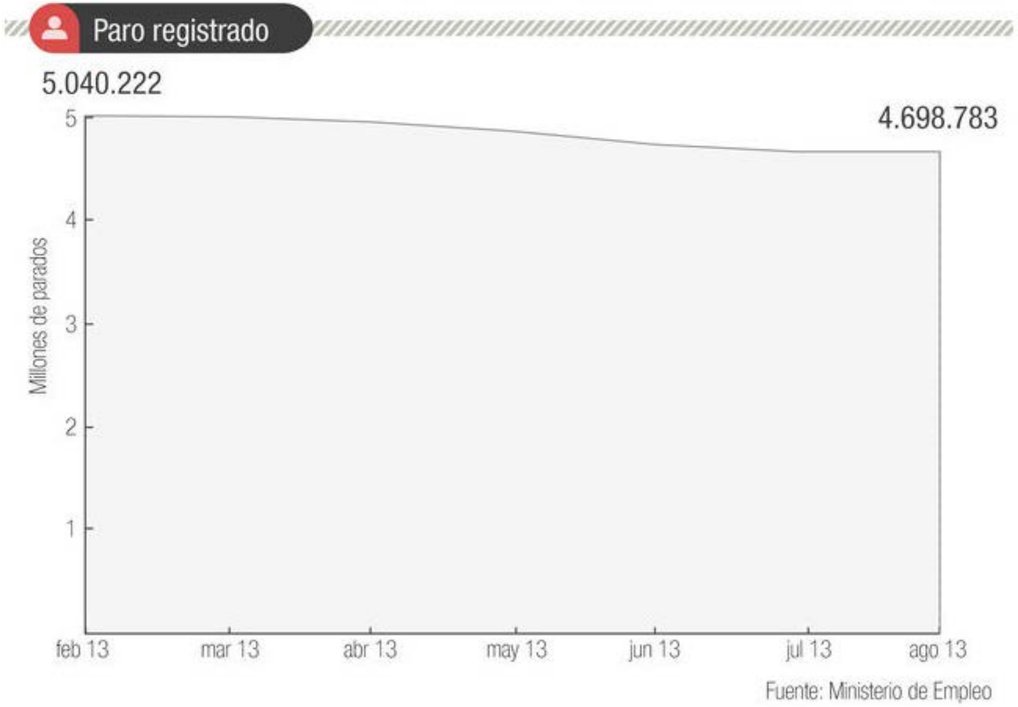
# Visual Honesty

# Some tips to make good graphs

- Pay attention to the axis
- Add annotations
- Show your data
- Avoid chart junk

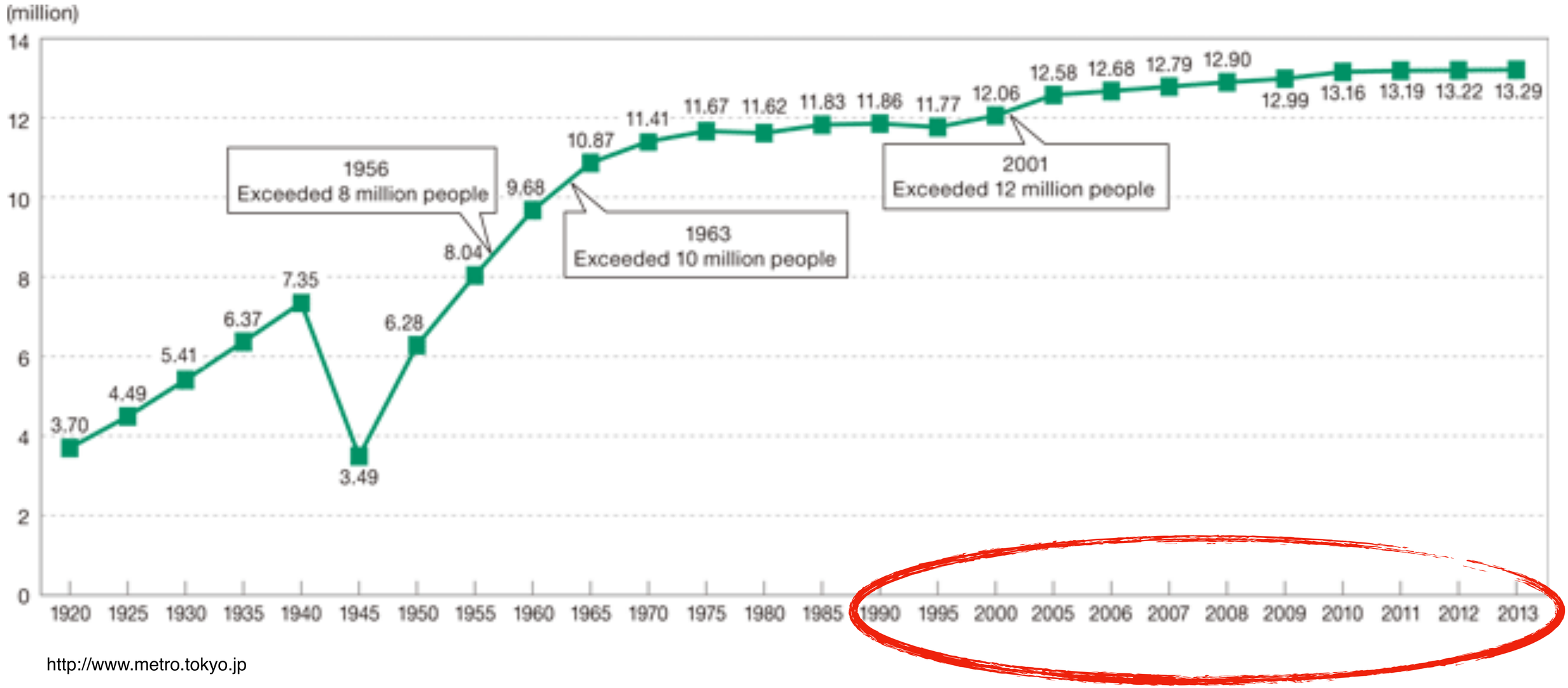
# Tips to make good graphs

- Pay attention to the axis
- Add annotations
- Show your data
- Avoid chart junk



Captura del gráfico con la evolución del paro registrado emitido en el Telediario 1 de TVE

# Tips to make good graphs




http://www.metro.tokyo.jp

# Tips to make good g

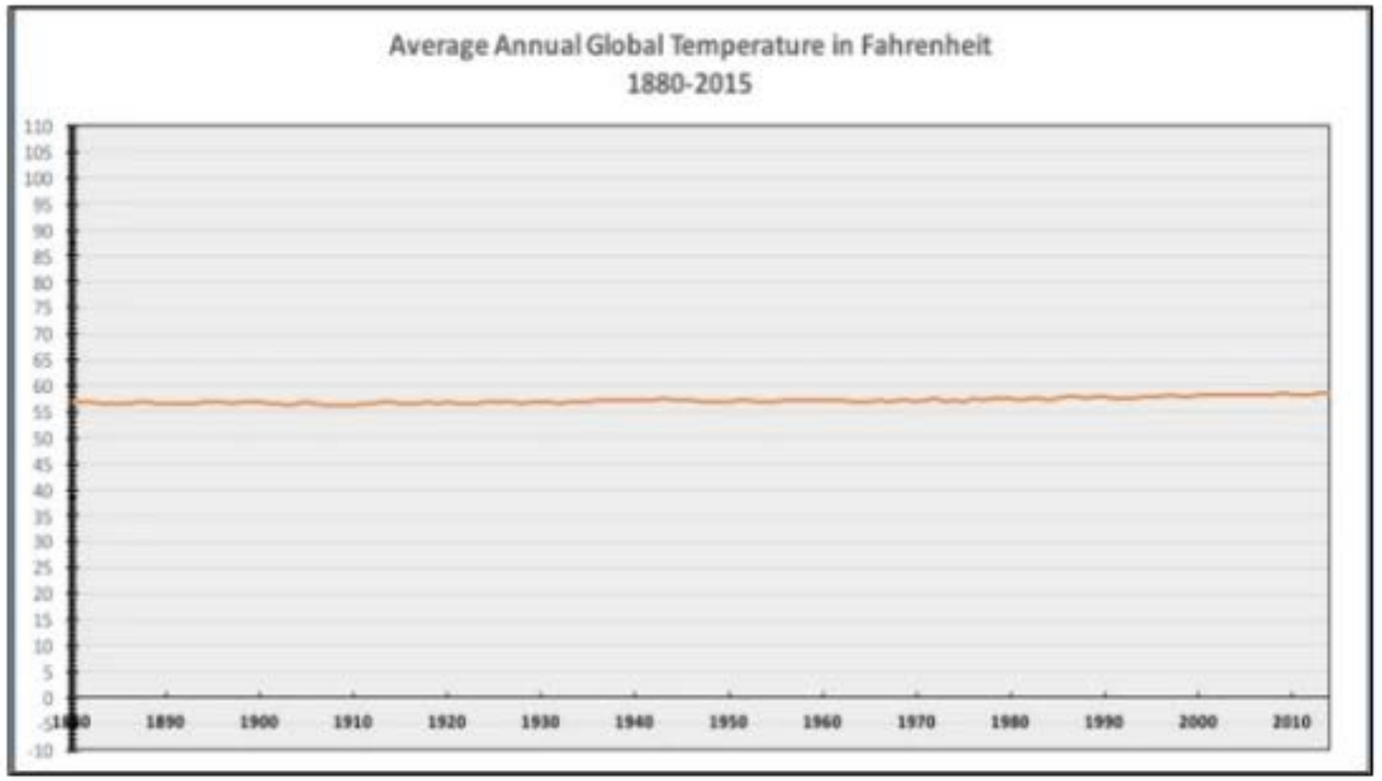
- Pay attention to the axis
- Add annotations
- Show your data
- Avoid chart junk

National Review
@NRO

Follow

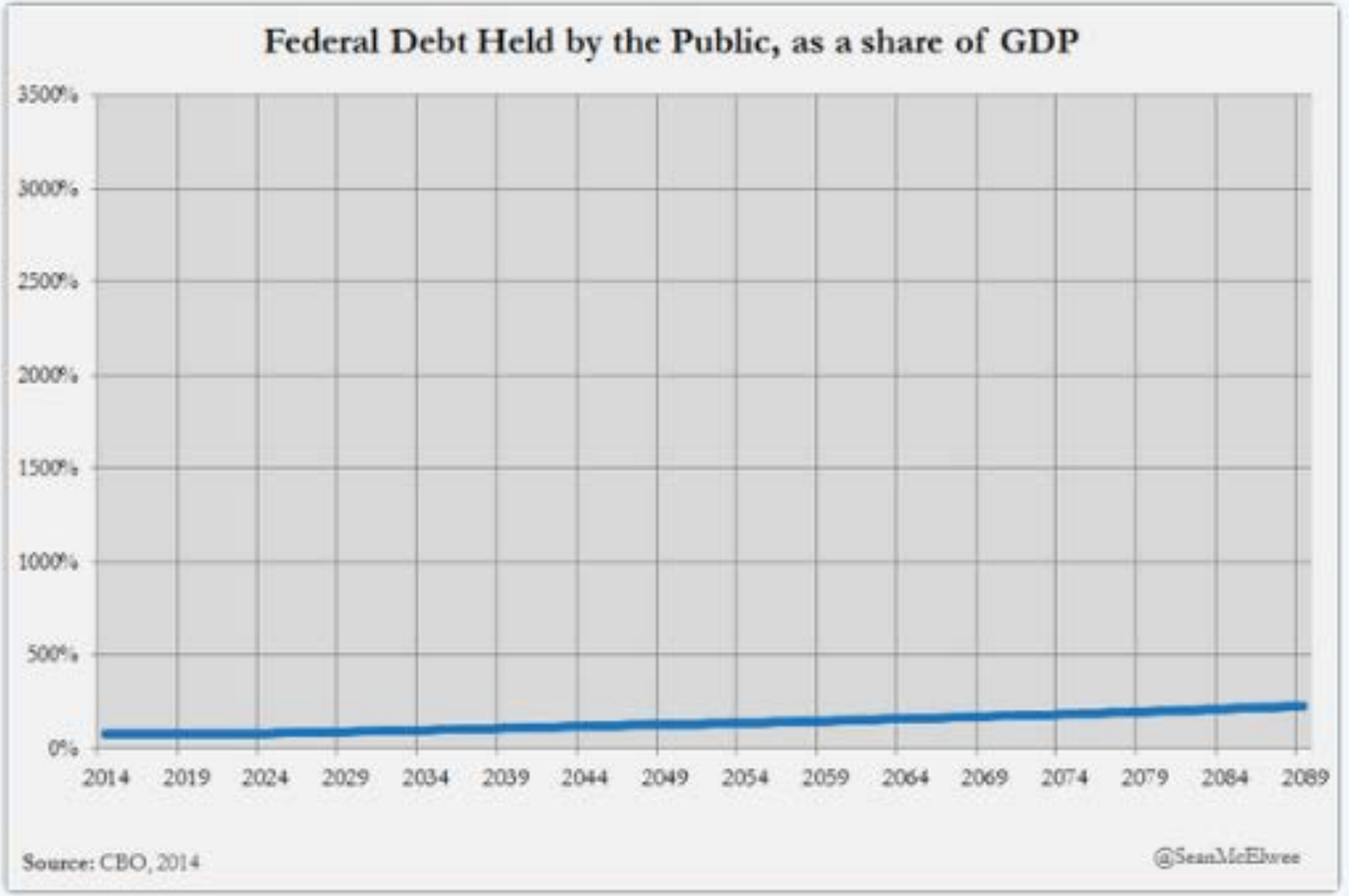
The only #climatechange chart you need to see.
natl.re/wPKpro

(h/t @powerlineUS)



**sean.** @SeanMcElwee · 14 Dec 2015

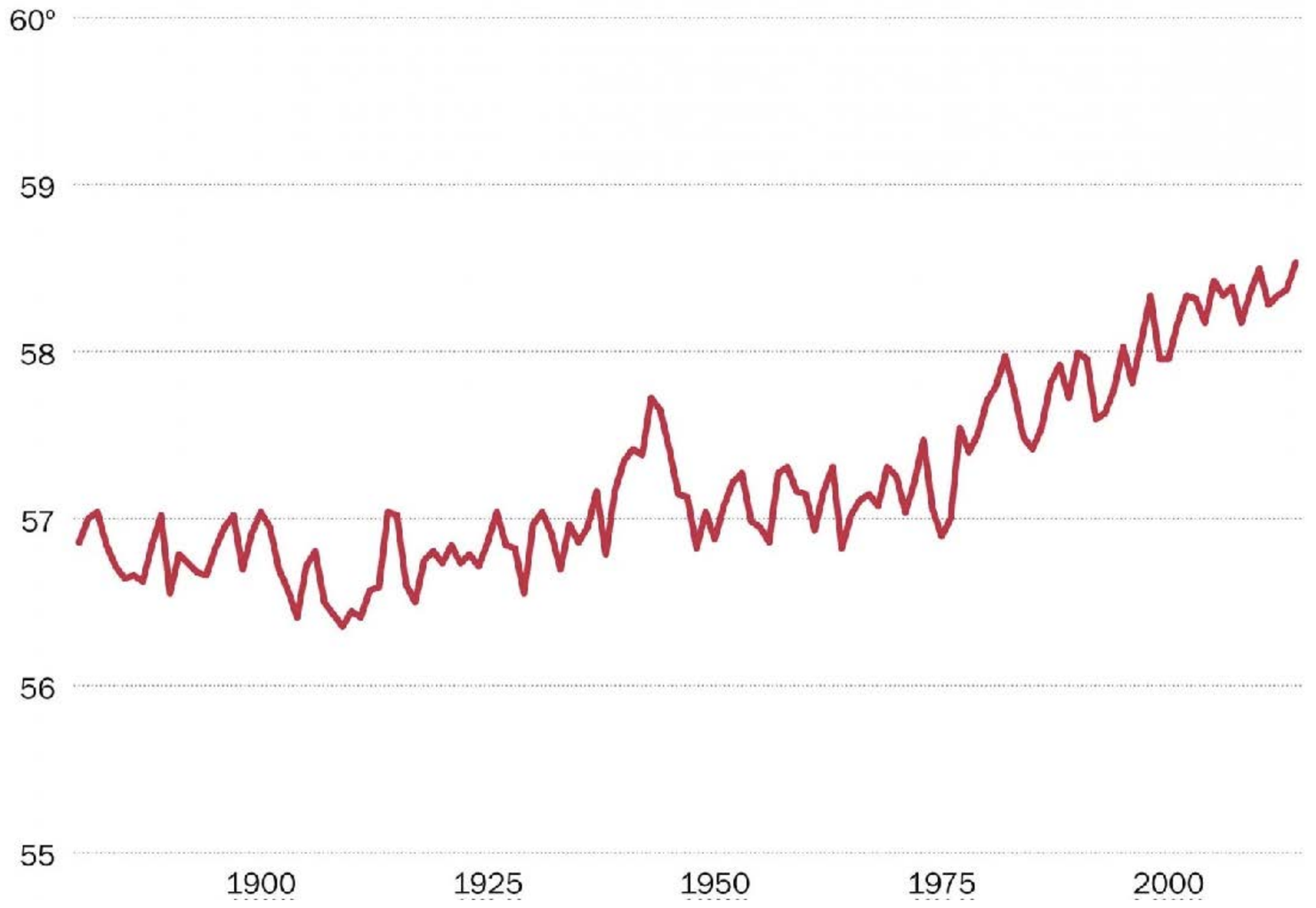
.@NRO @powerlineUS no need to worry about the national debt then either!




467 1.7K

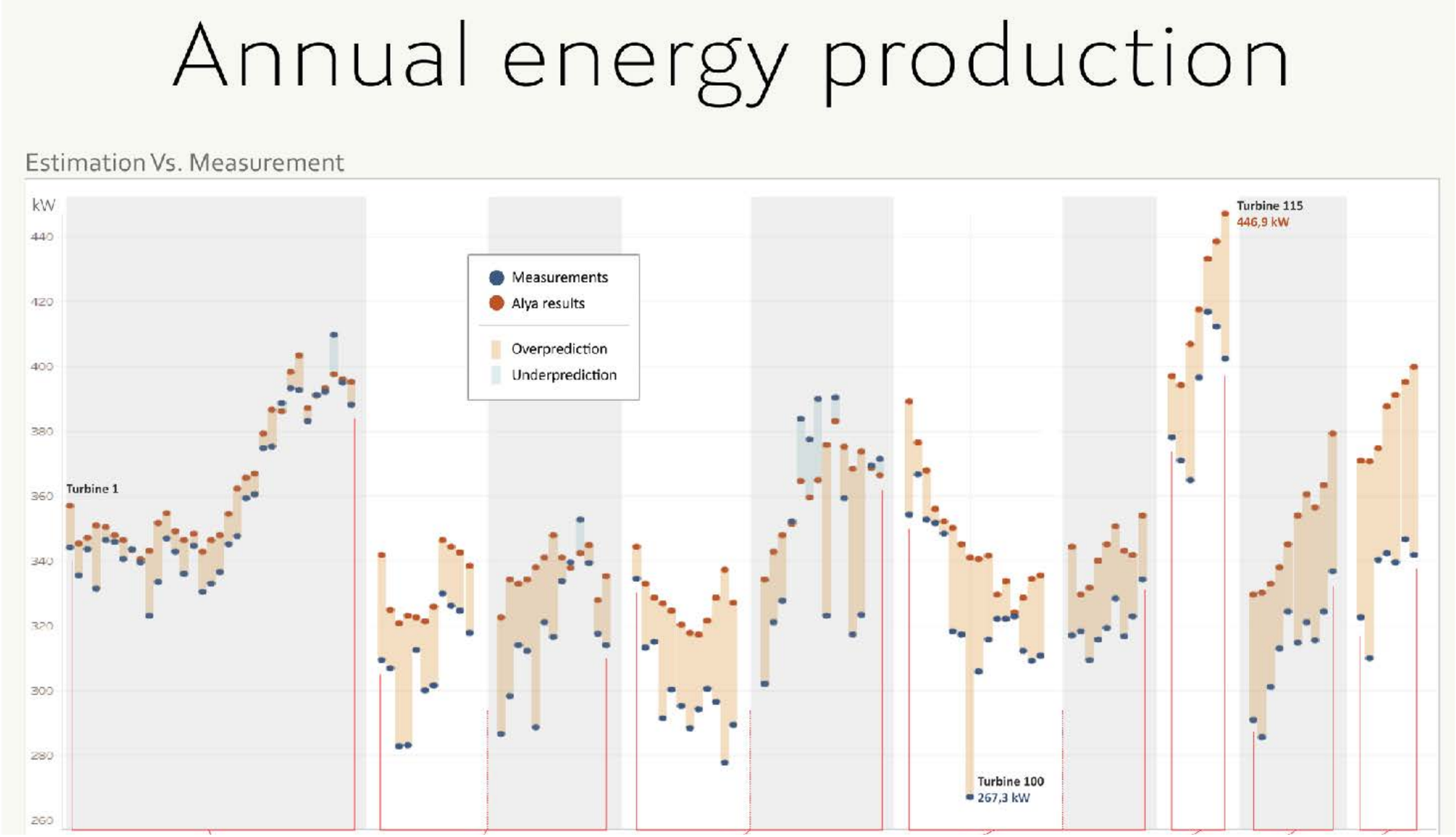
# Average global temperature by year

Data from NASA/GISS.




The Washington Post

Energy production of a wind farm in kW. We are comparing values to each other, don't need to start at 0

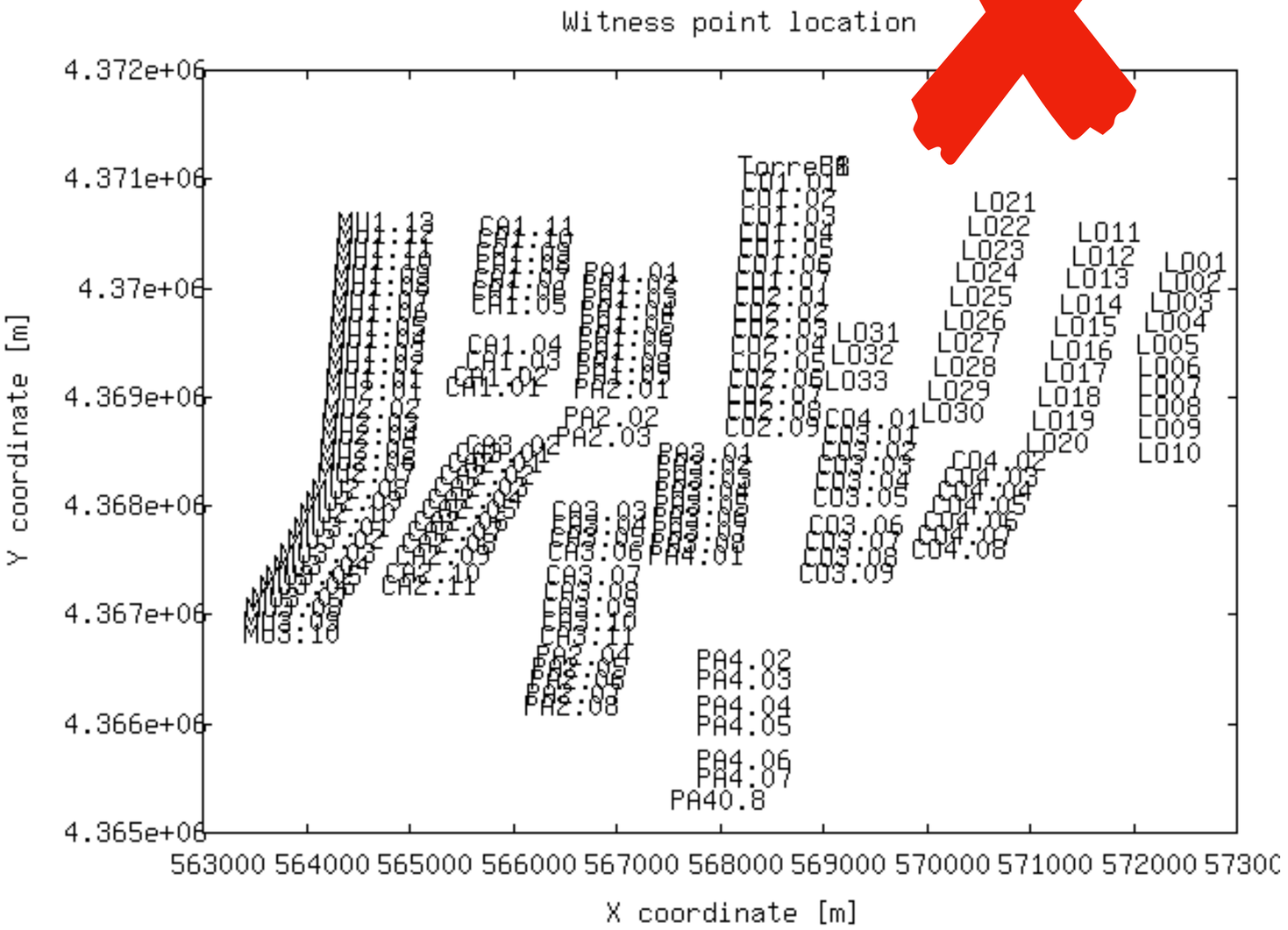



BSC Viz Team

# Tips to make good graphs

- Pay attention to the axis
- Add annotations
- Show your data
- Avoid chart junk

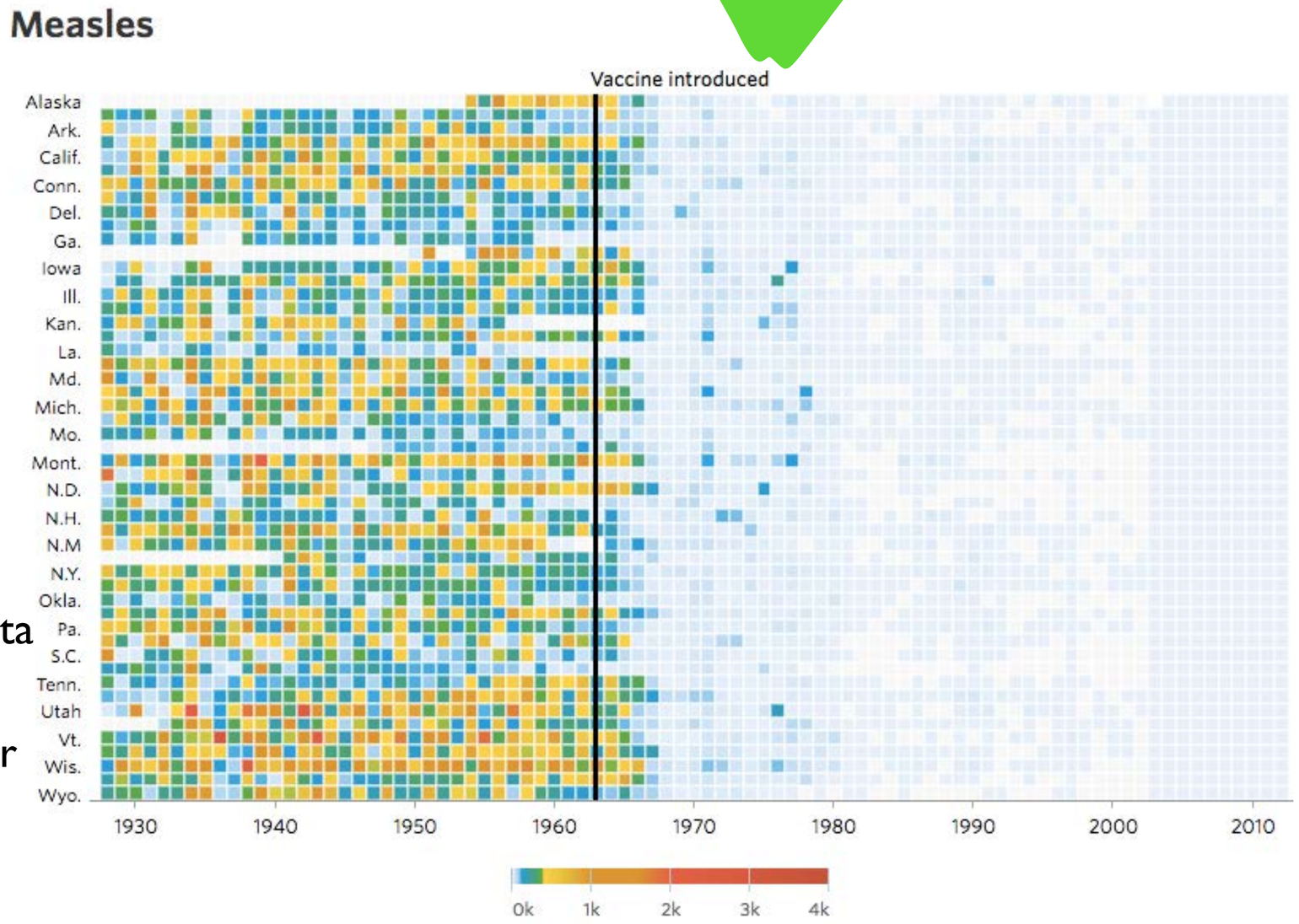
Be careful with default plot settings:
Put clear marks on the axis, understandable text/data points, meaningful titles



# Tips to make good graphs

- Pay attention to the axis
- Add annotations
- Show your data
- Avoid chart junk

Don’t assume your audience knows your data
**They don’t!**
State axis clearly, put titles and legends, color scale, highlight important events, etc.

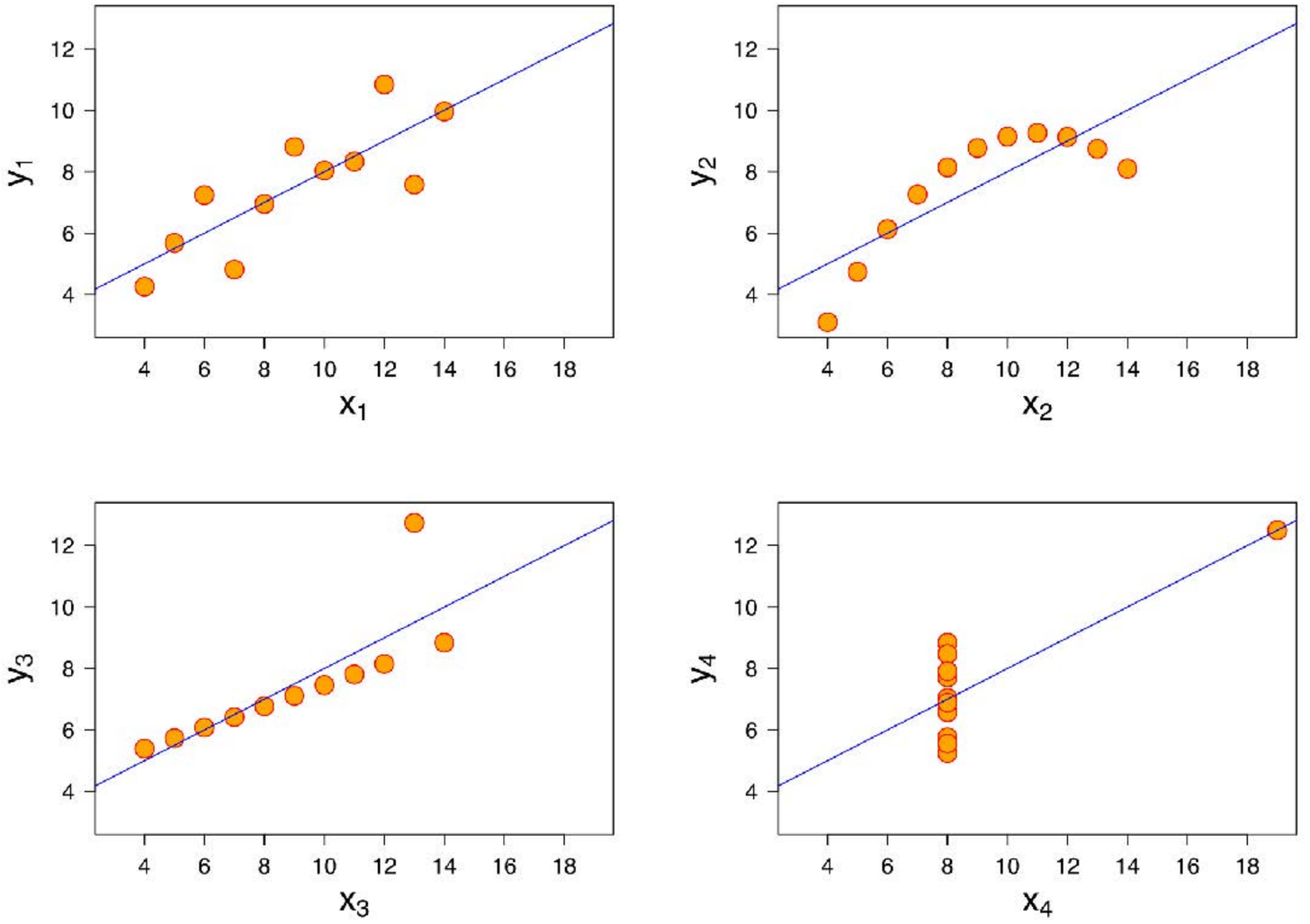


# Tips to make good graphs

- Pay attention to the axis
- Add annotations
- Show your data
- Avoid chart junk

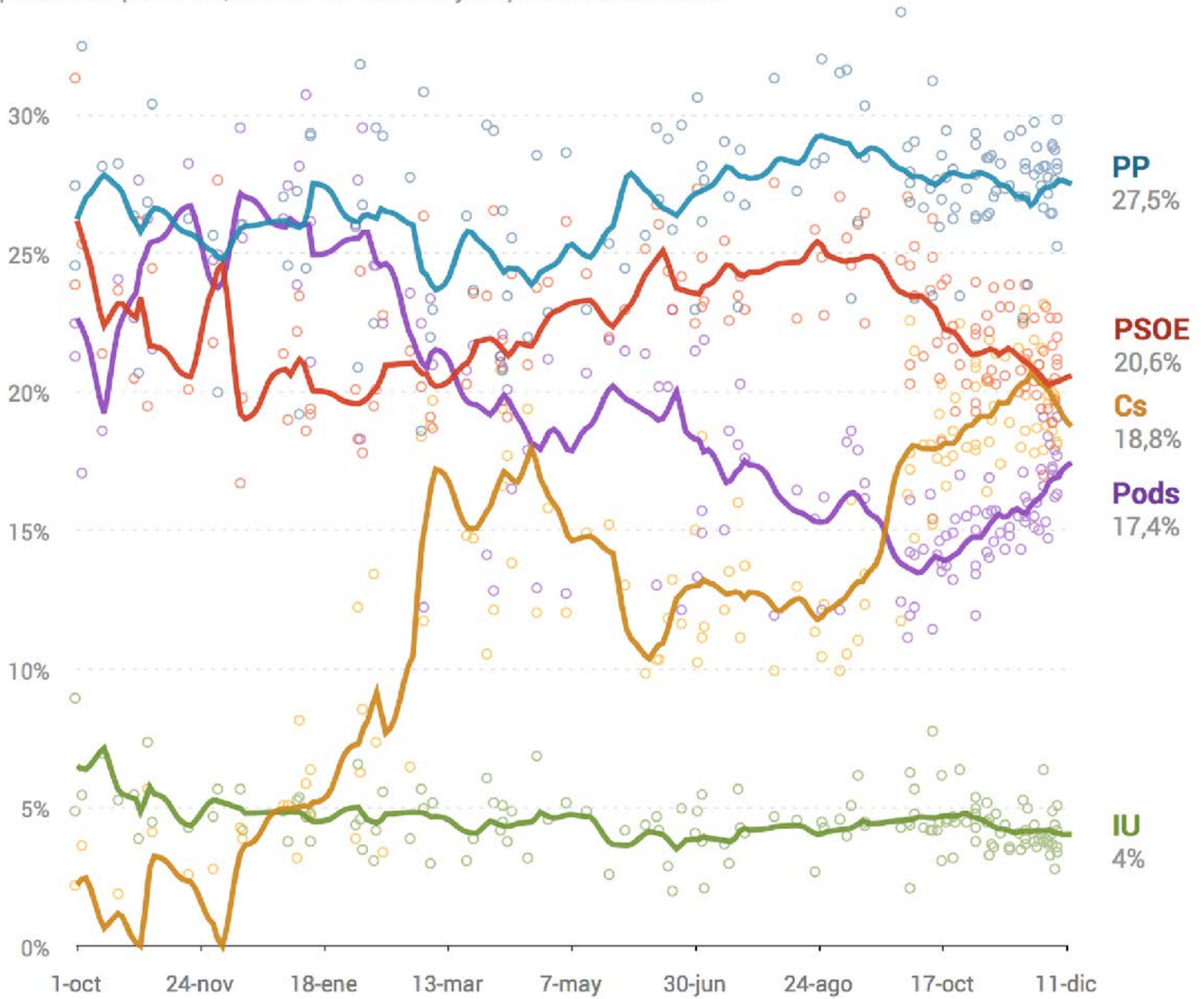
Anscombe's quartet, different distributions with identical mean, median, and other summary statistics

The shape of the distribution can reveal more about the data




from Wikipedia

Porcentaje de voto según las encuestas. Las líneas representan un promedio ponderado por fecha, tamaño de muestra y empresa encuestadora.



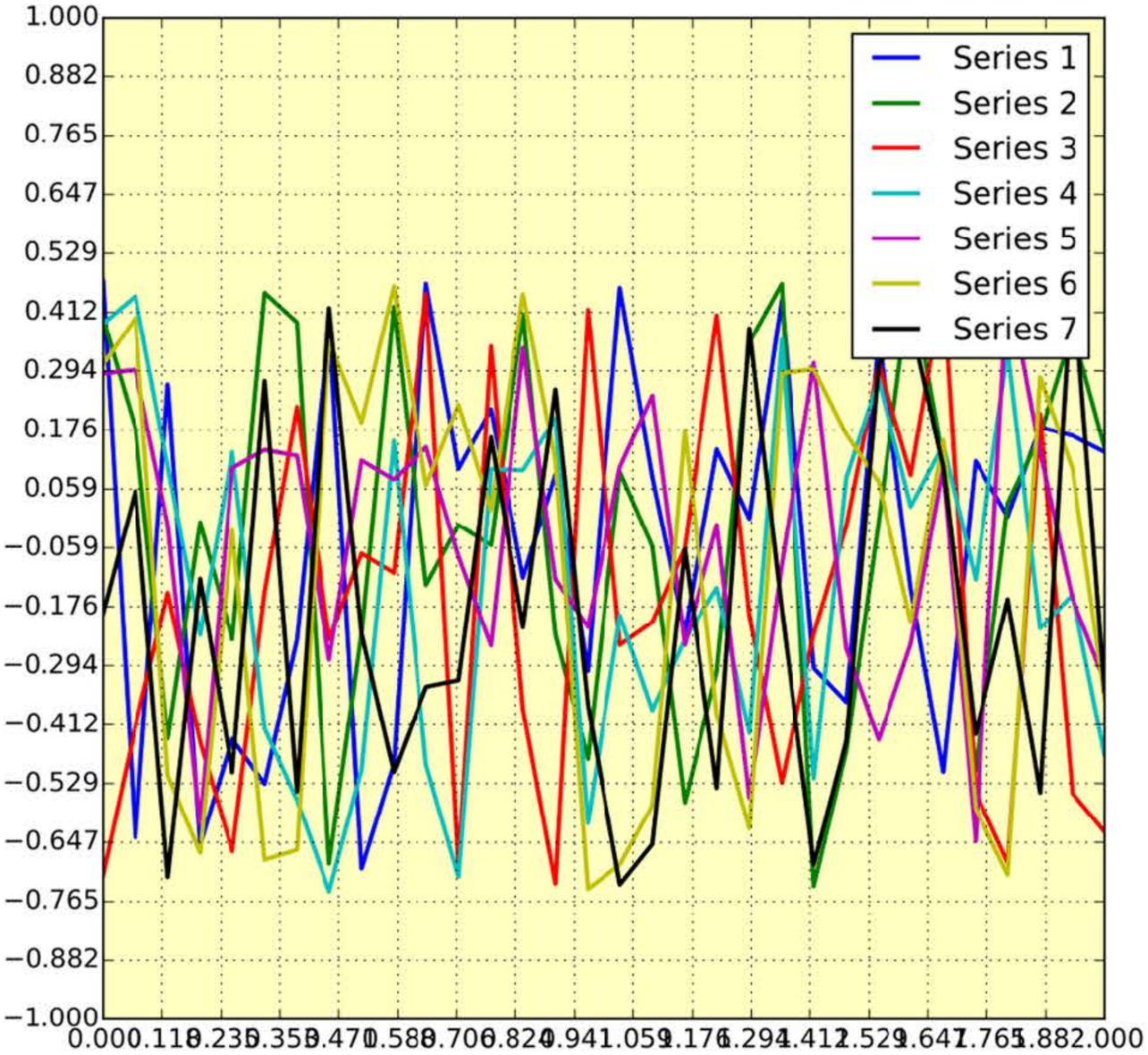

Fuente: Electograph. Fecha actualización: 14 dic.

Kiko Llaneras | EL ESPAÑOL

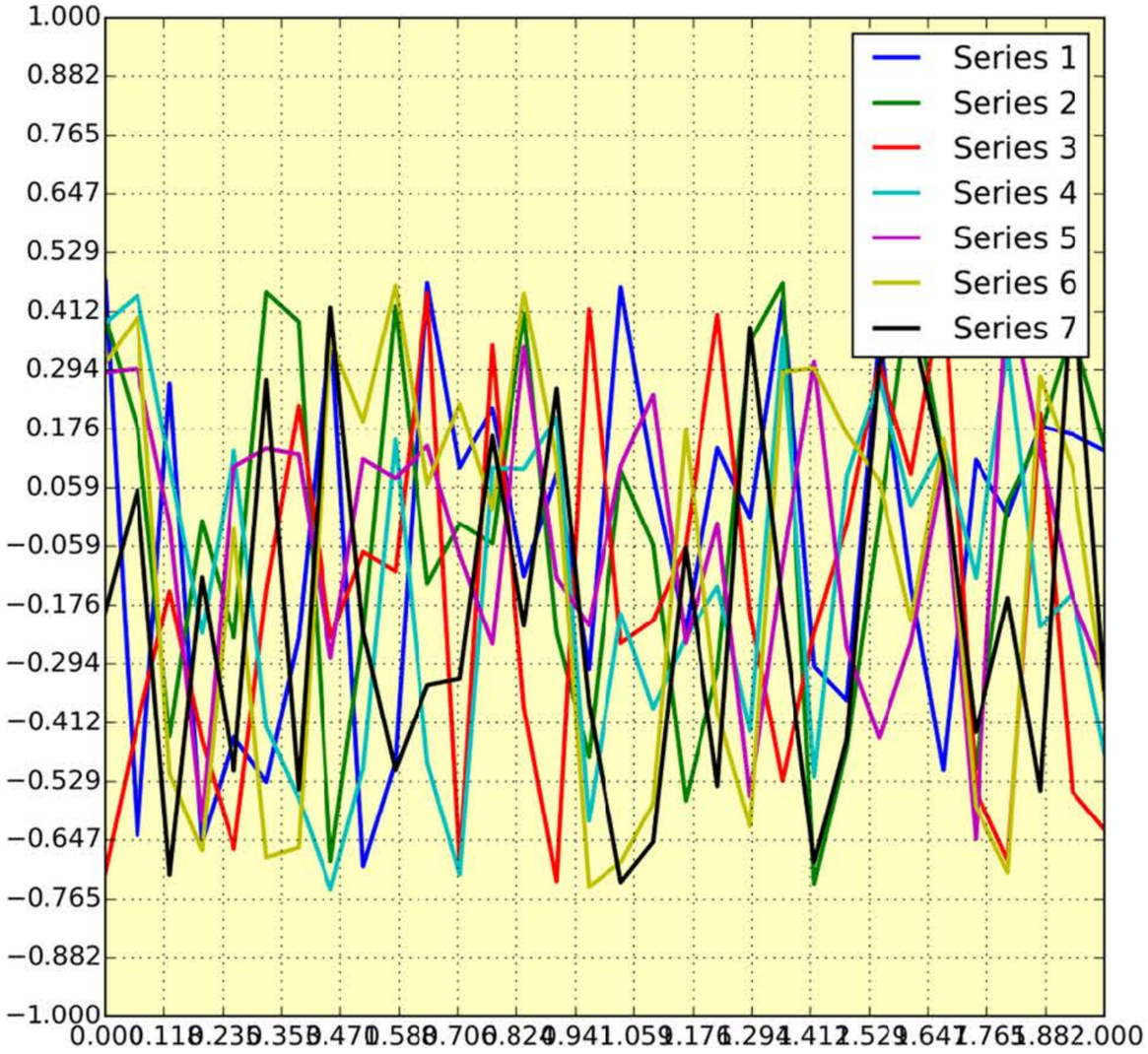
# Tips to make good graphs

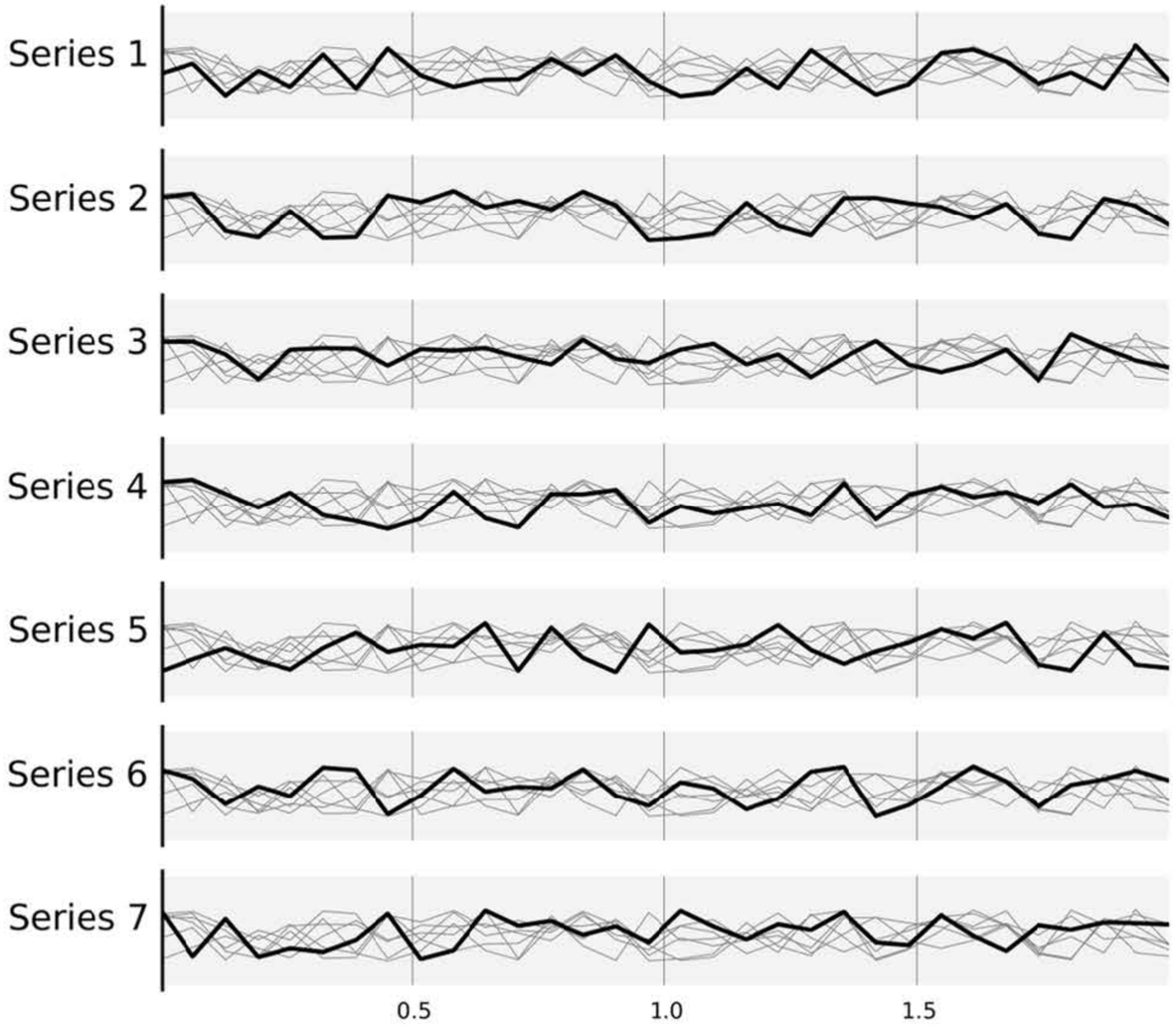
- Pay attention to the axis
- Add annotations
- Show your data
- Avoid chart junk

Avoid over-charging your charts:
Less colours, transparent/neutral background, meaningful dimensions and annotations, unobtrusive legends, etc.



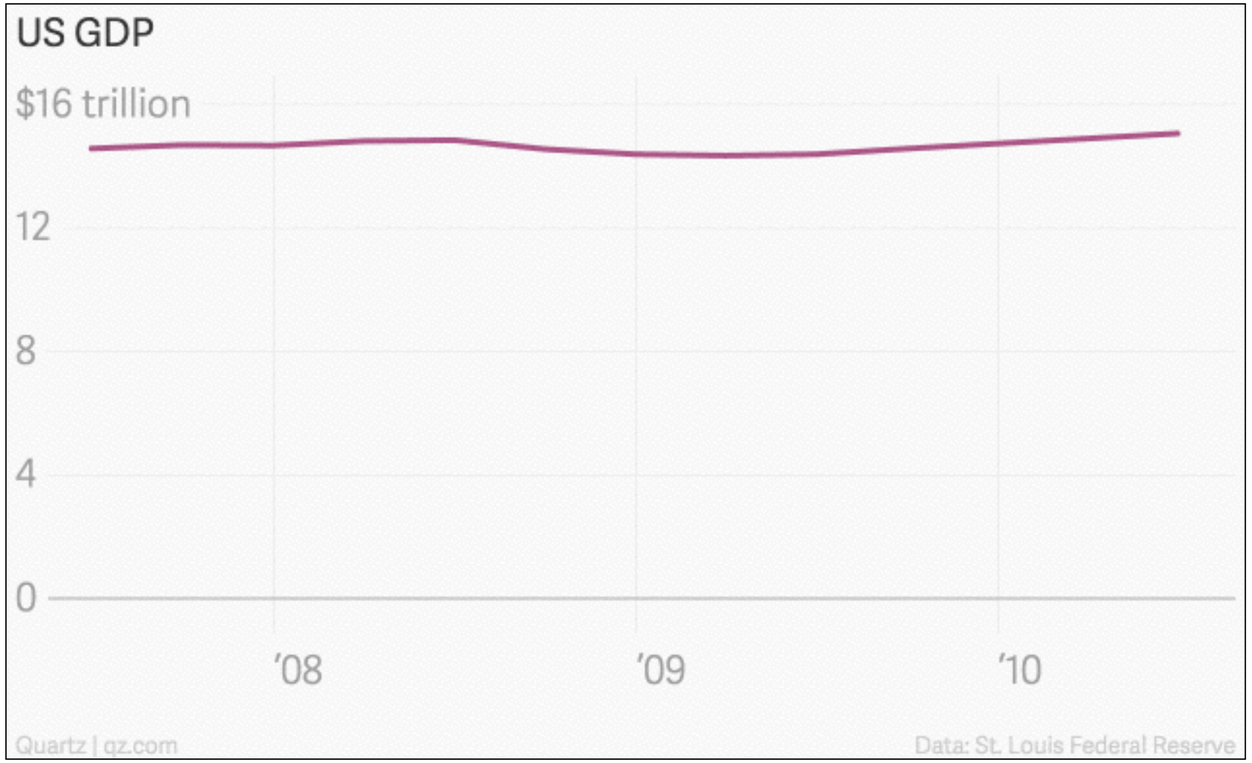
For the objective of comparing trends, split in multiple linecharts work better than one

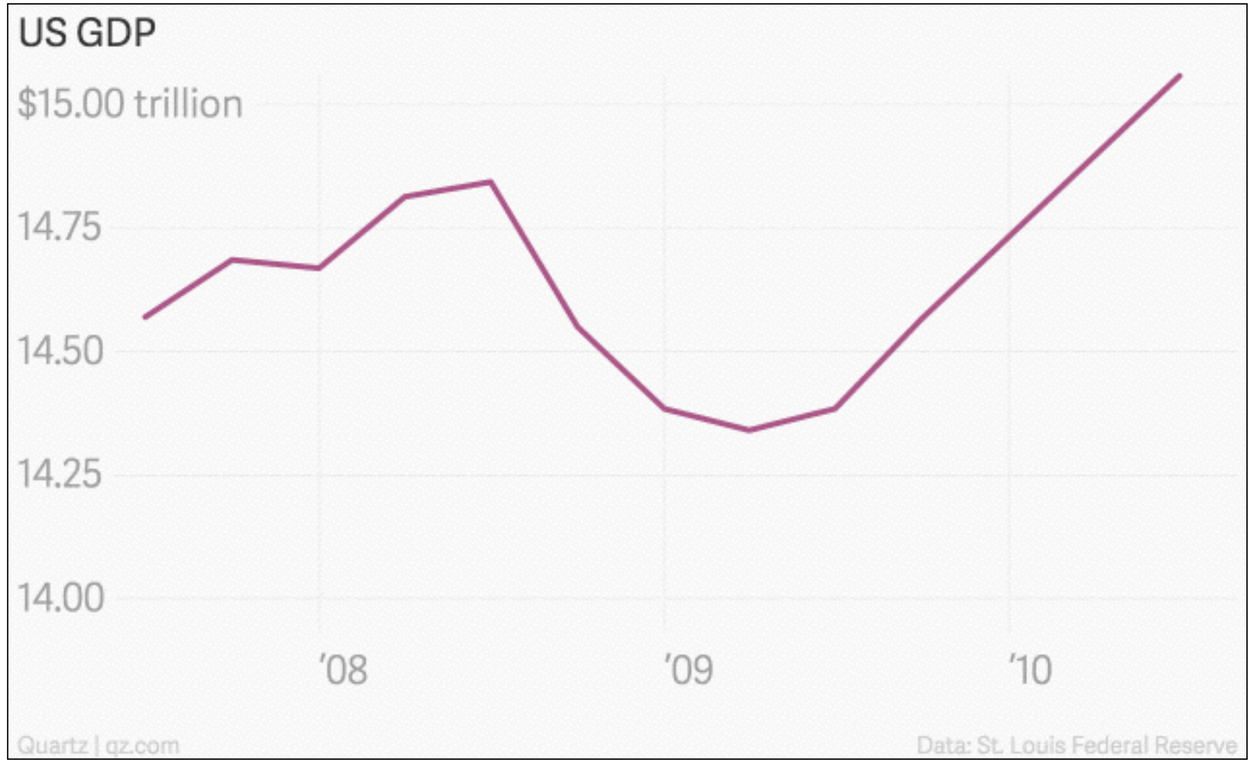


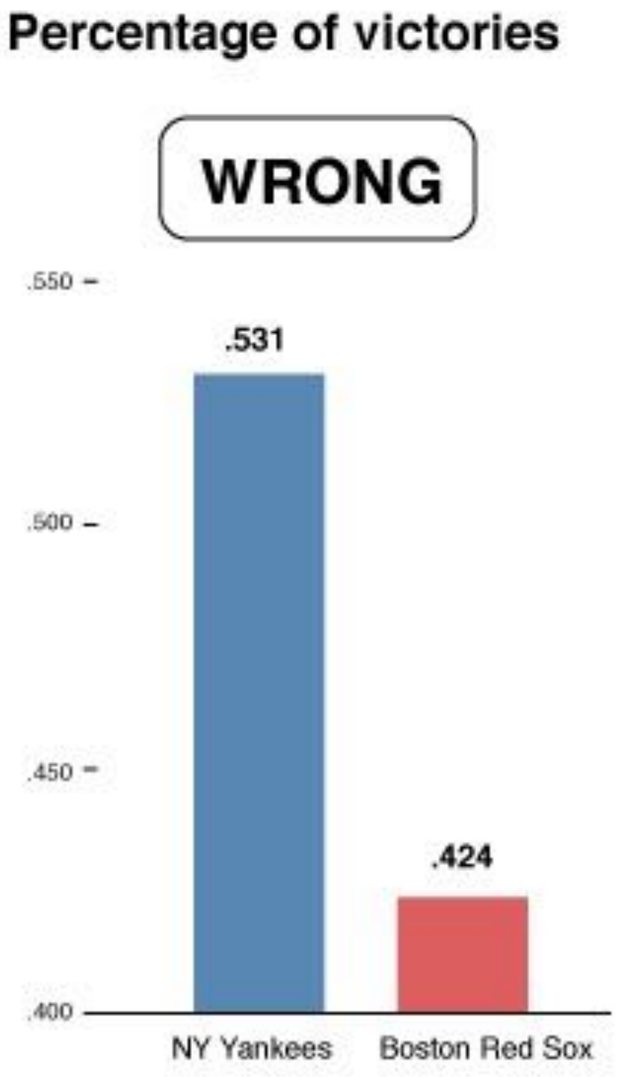
# Visual precision

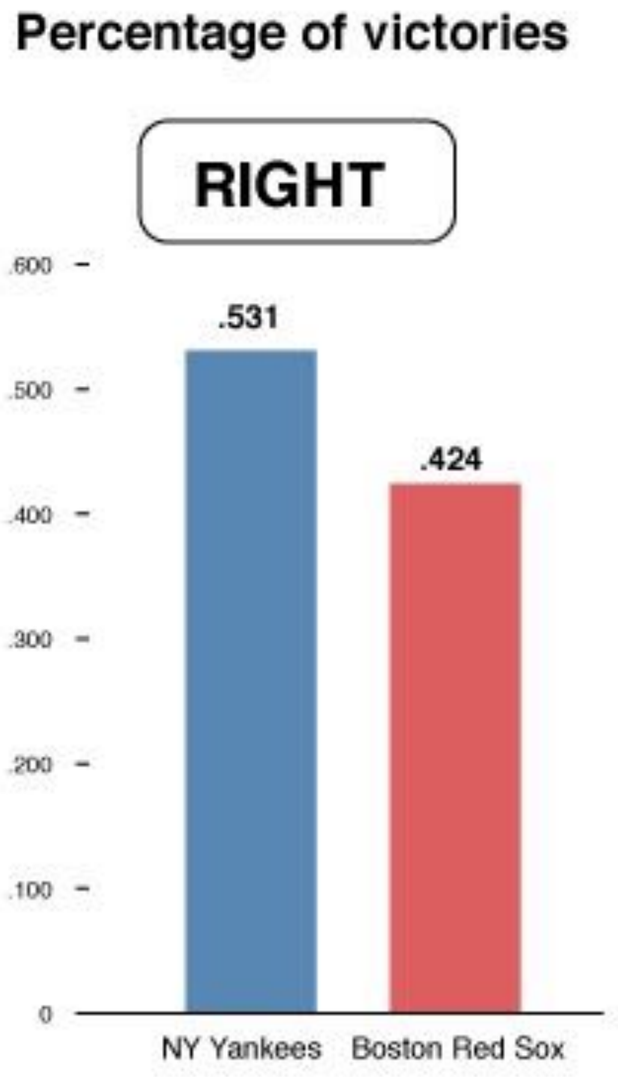
# Choice of axes: highlight trend or variation



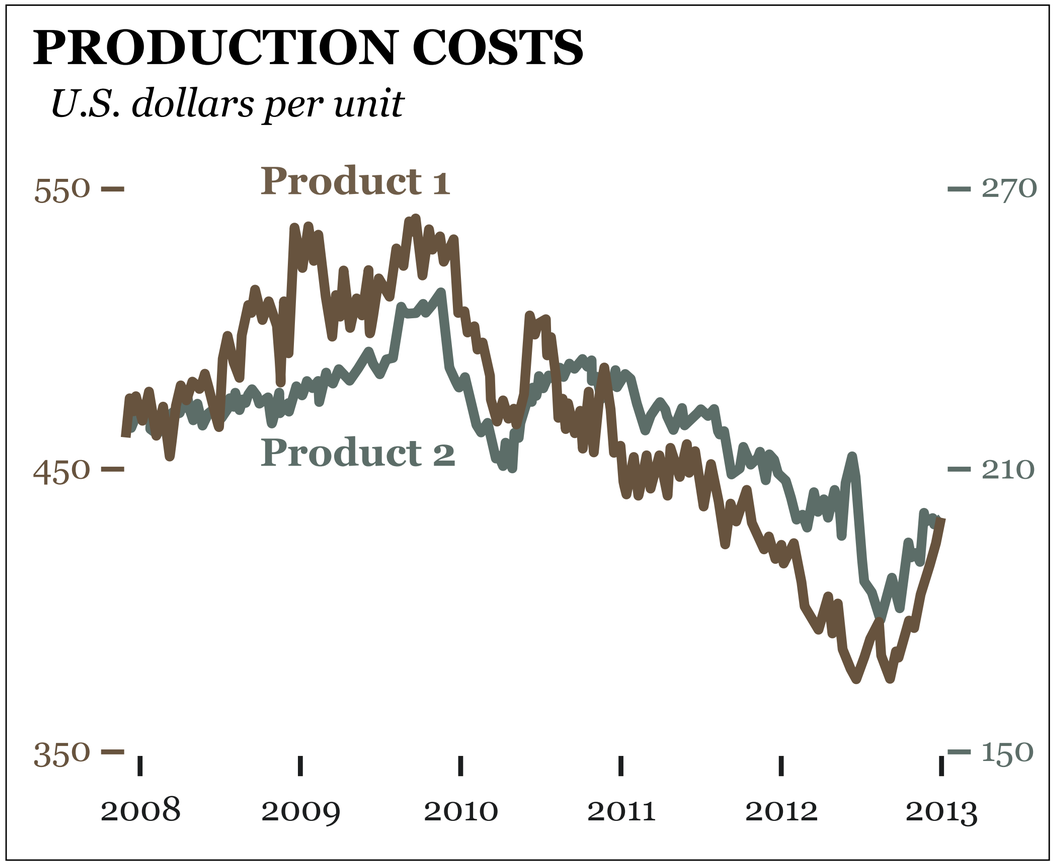
# Don’t cut the Y-axis in percentages

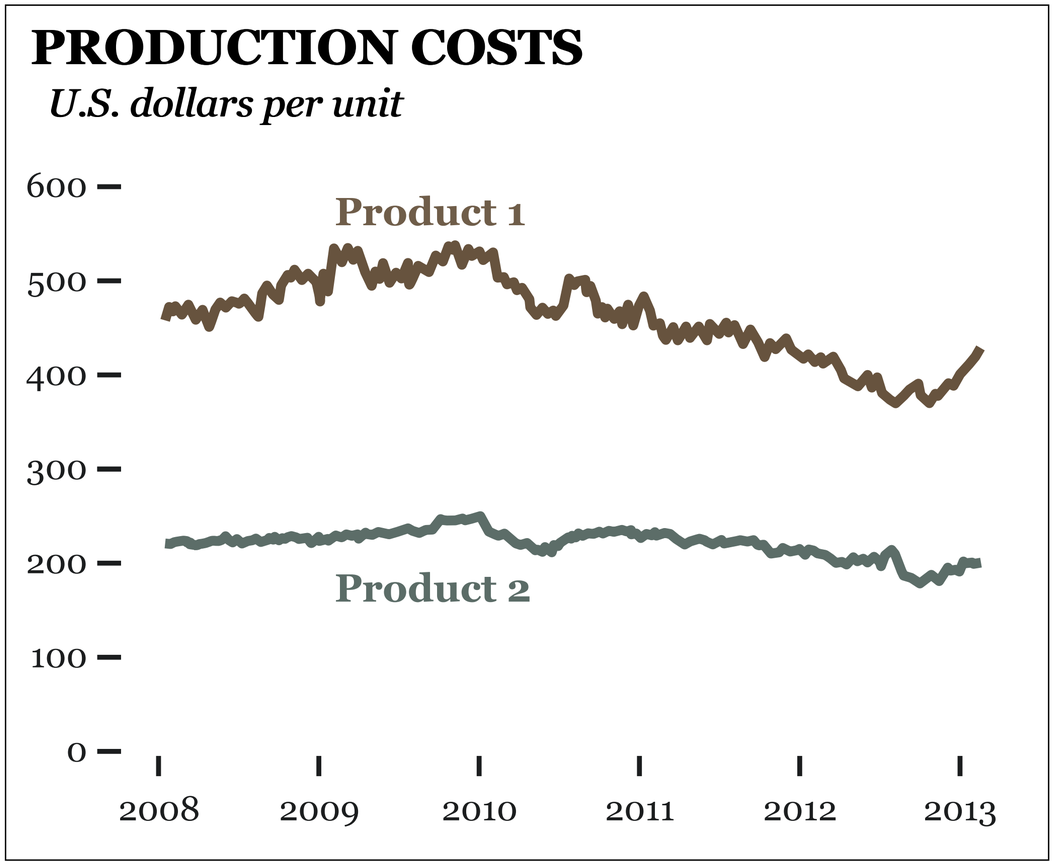


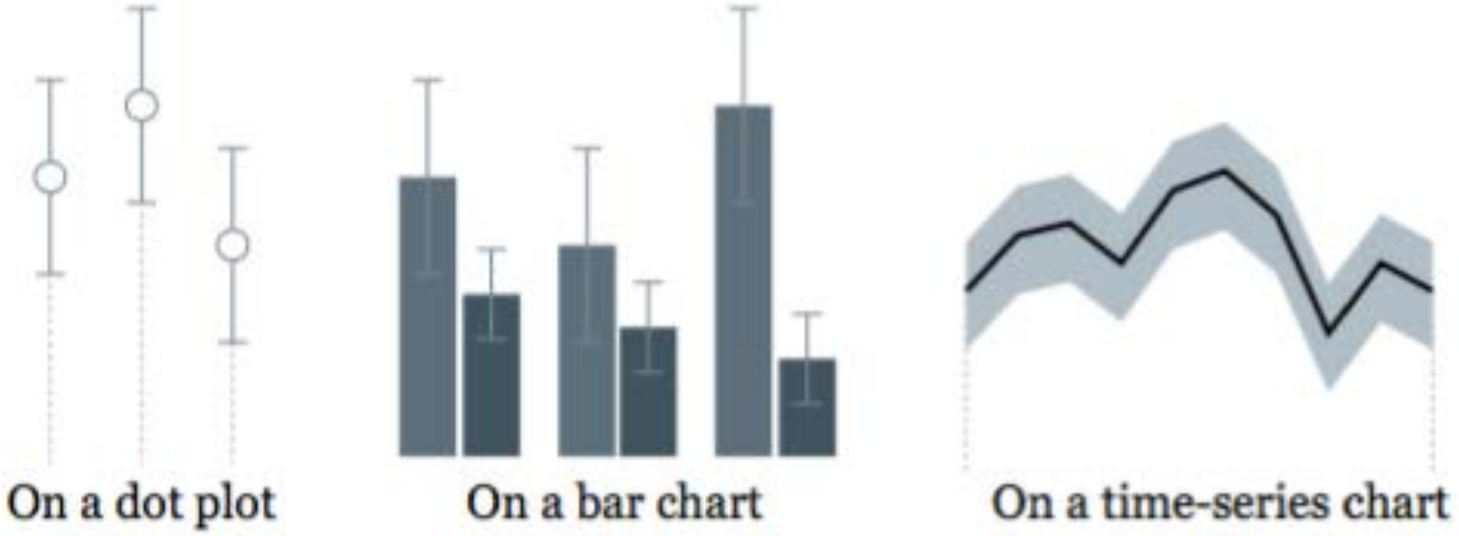

http://news.nationalgeographic.com/2015/06/150619-data-points-five-ways-to-lie-with-charts/

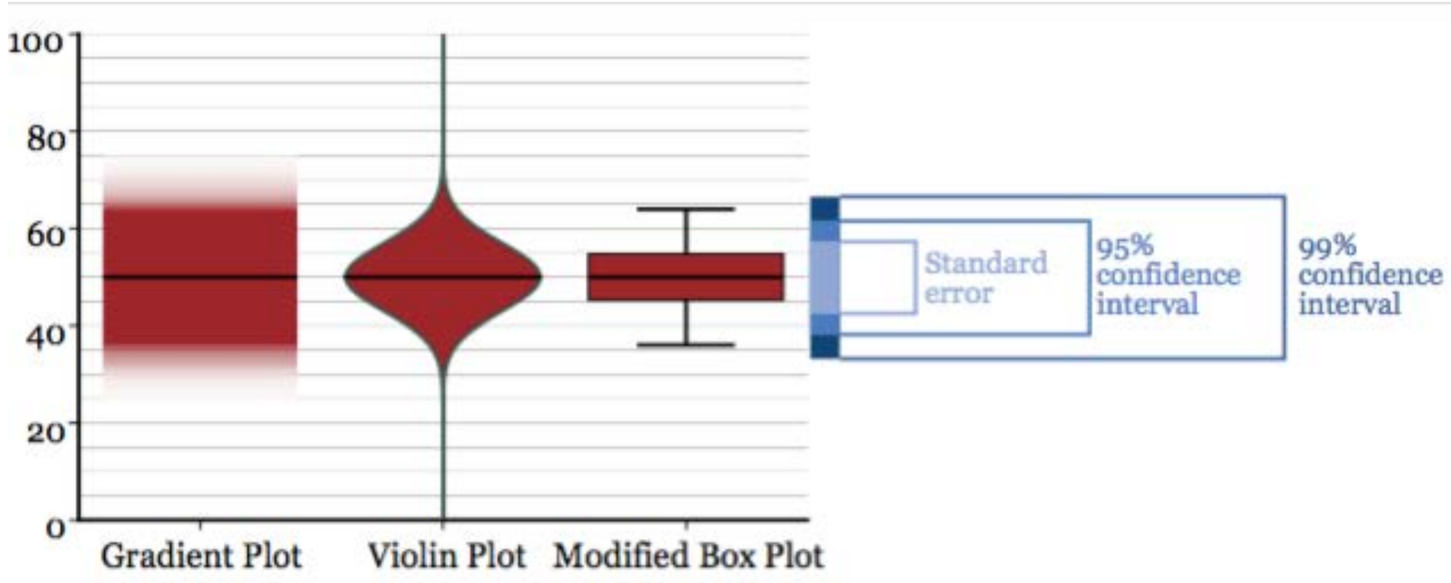
# Be careful with double axes - False perception of equivalence



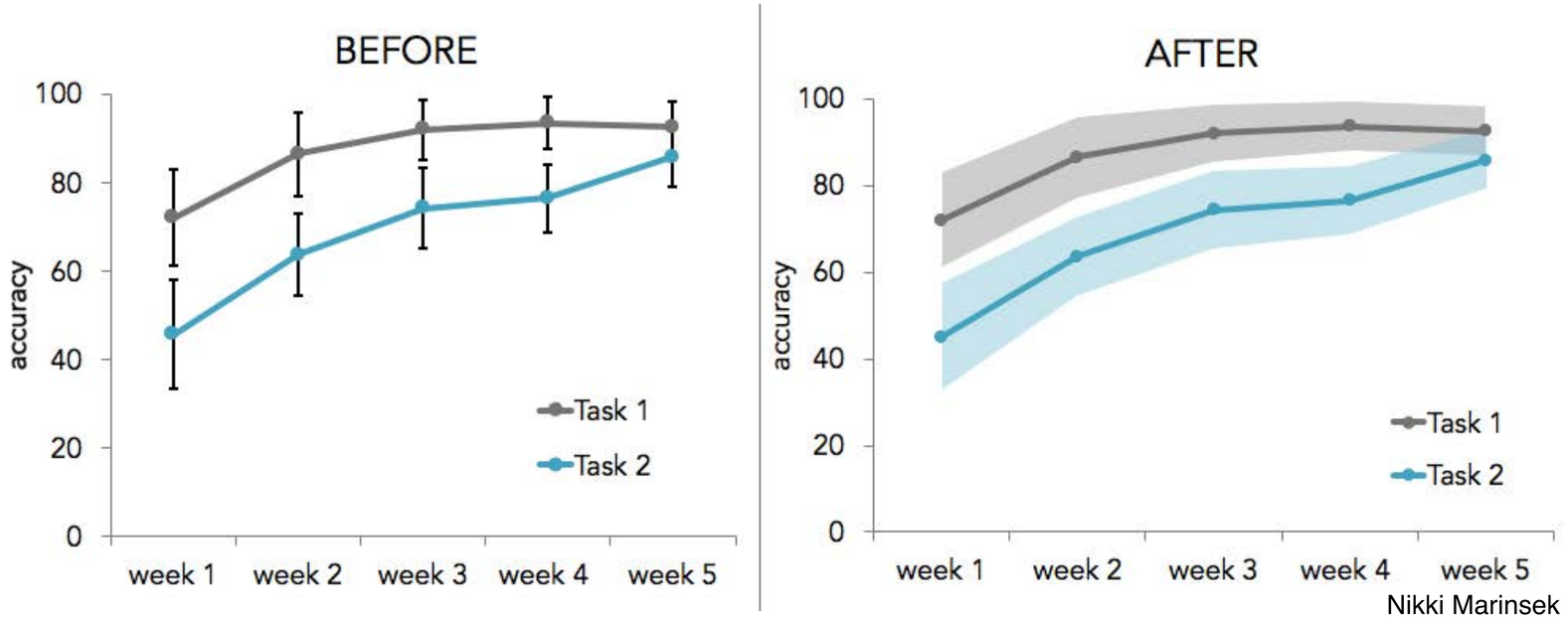



# Uncertainty visualisation



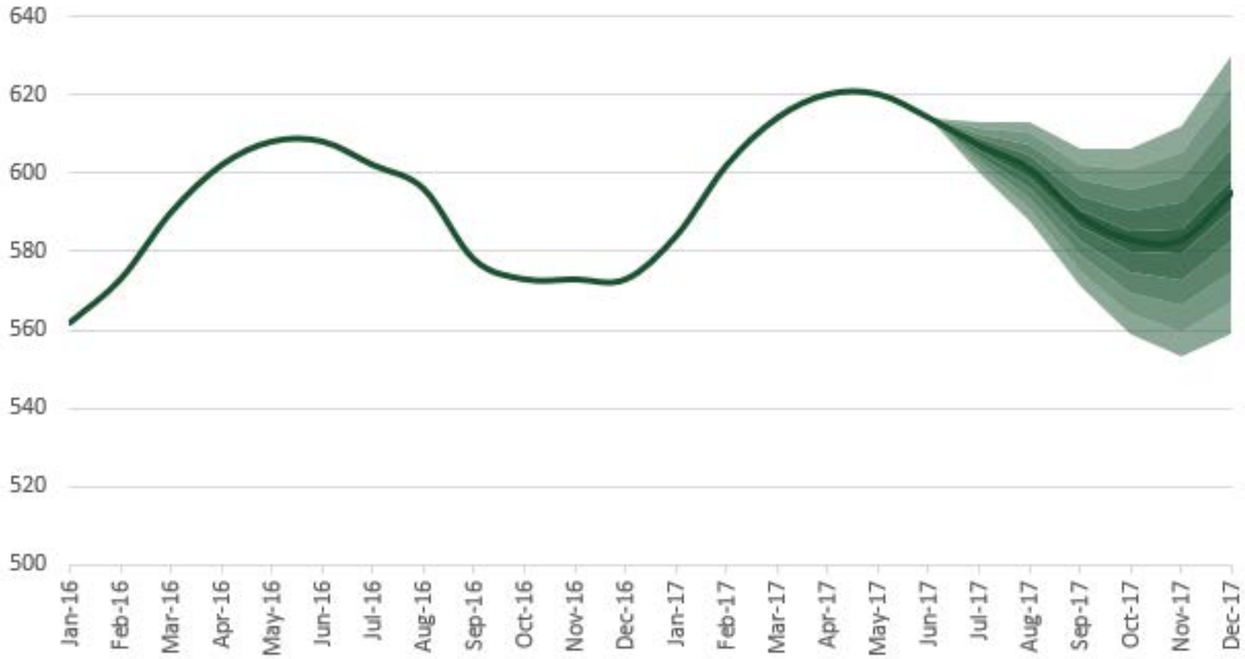




Error bars considered harmful. Correll M., Gleicher M.




Nikki Marinsek

# Uncertainty visualisation




exceloffthegrid.com

Choose the right plot and avoid distorting the data

# Now what?

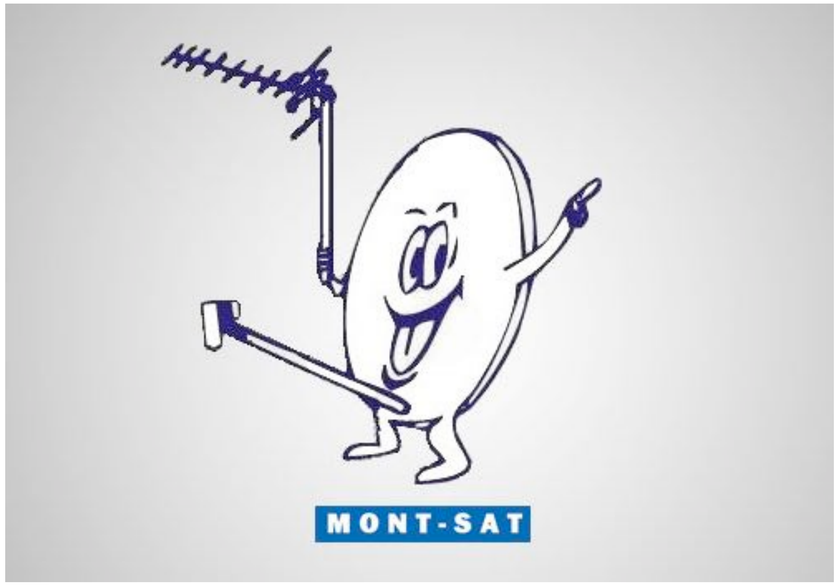
Graphic Design

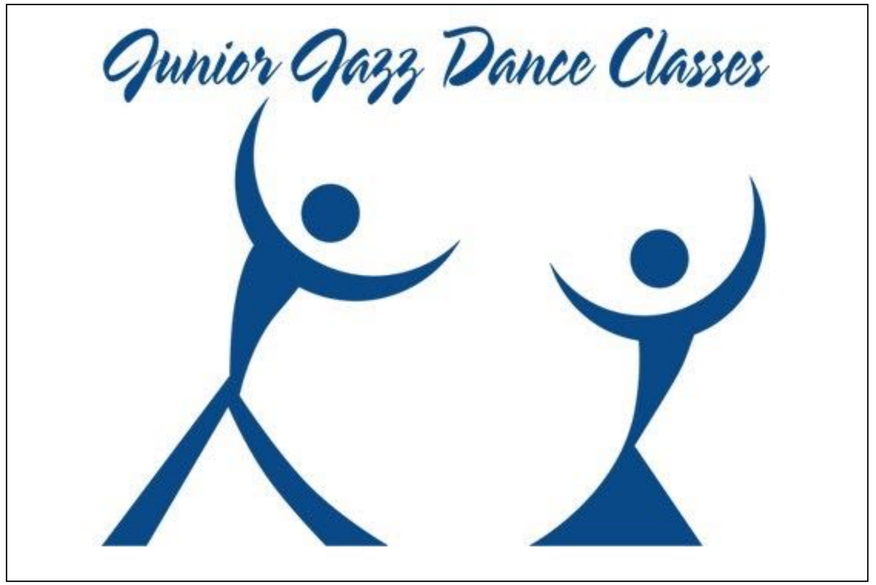
A quick word on **design**

# Not just about aesthetics

# Not just about aesthetics

## It has to work!



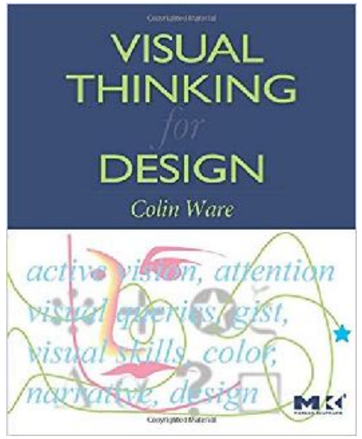



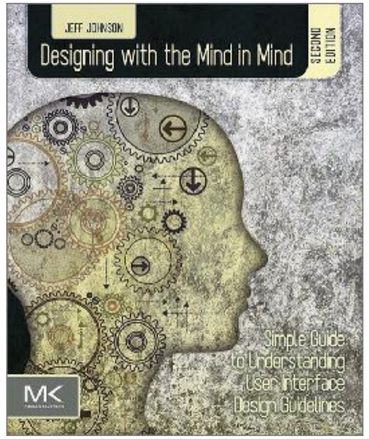
# Not just about aesthetics

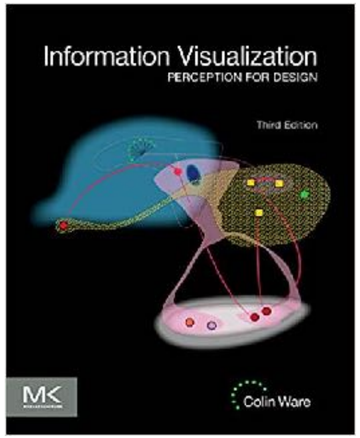
## It has to work!

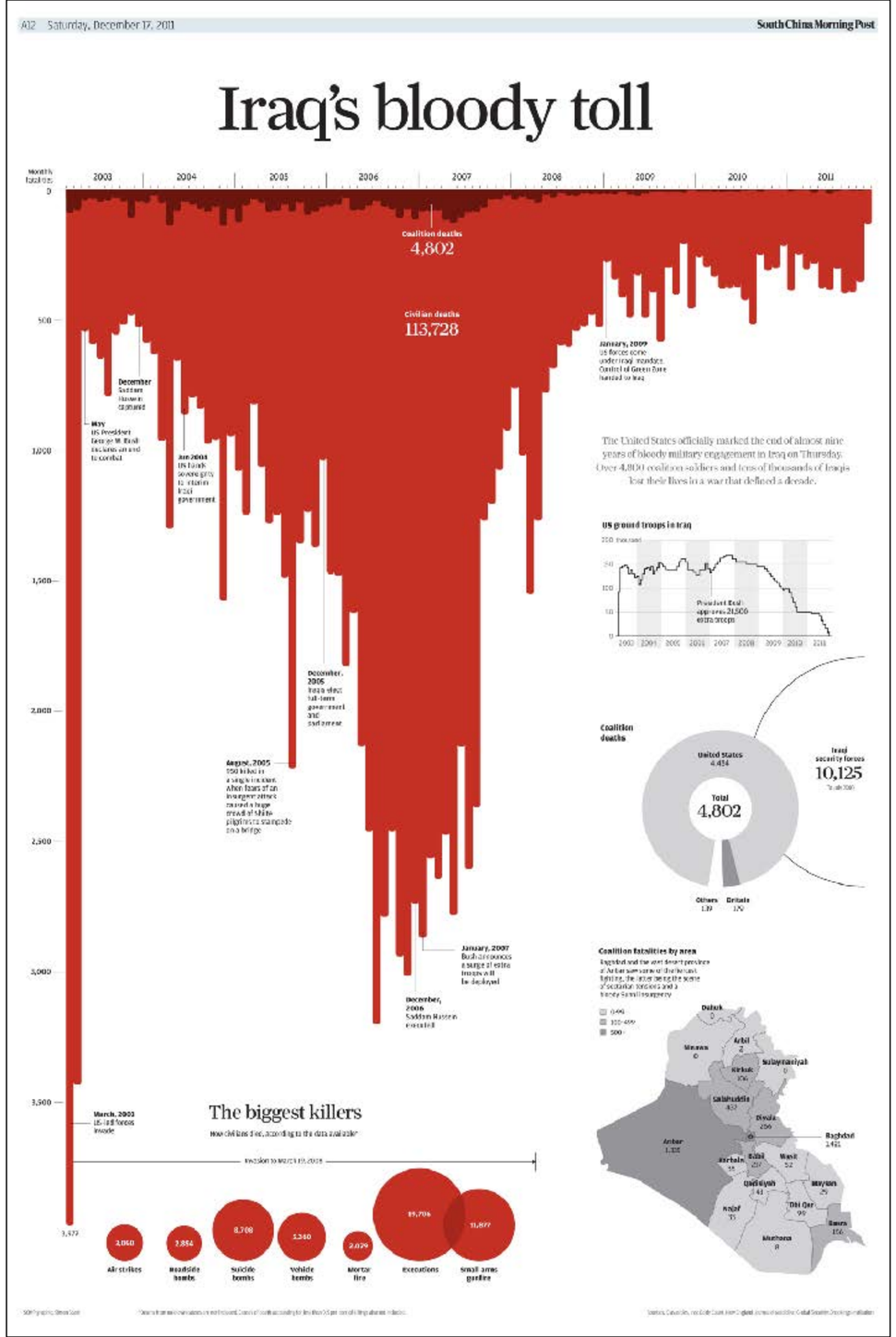
# Active Vision

- Graphic designs are cognitive tools that facilitate insight
- Vision and cognition are tightly related
- Puts human perception at the center of design: Design with the mind in mind



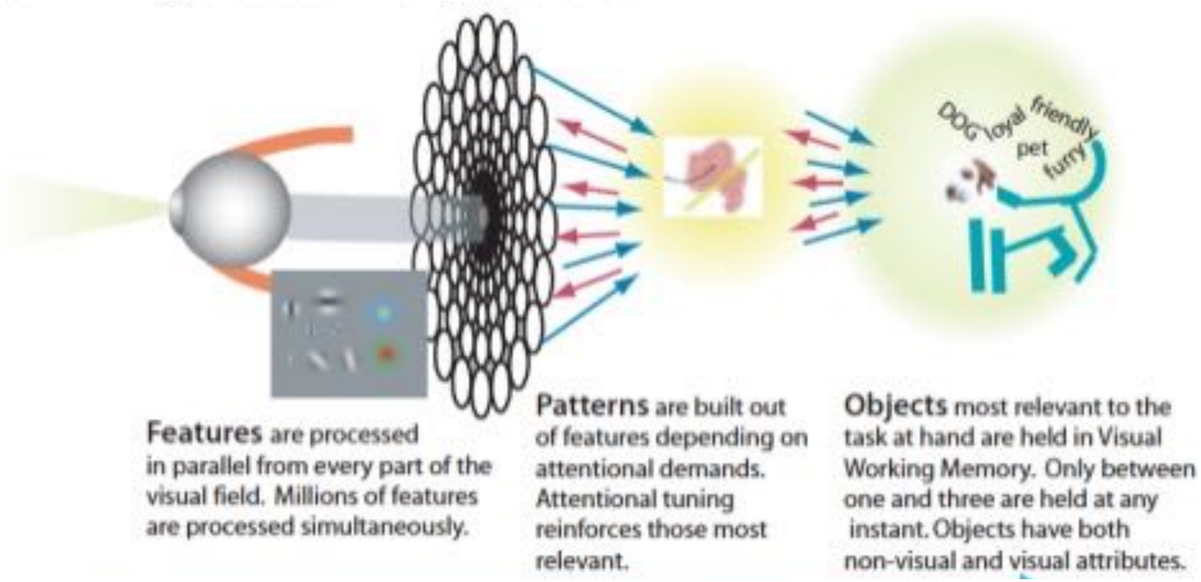



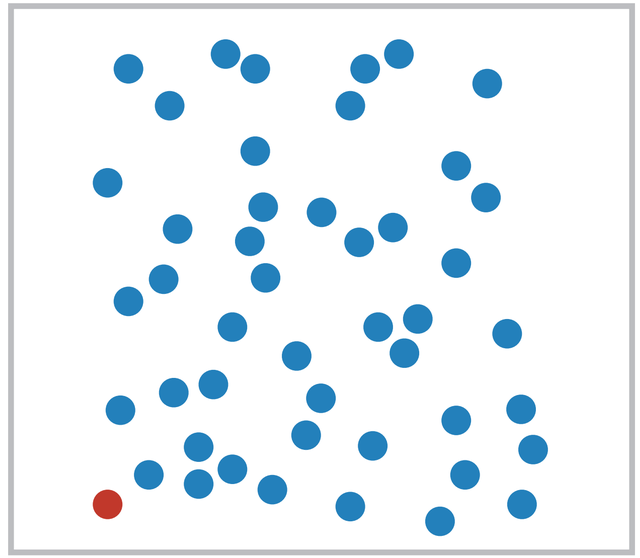
How many passes does the team in white make?

# Visual perception

- We don’t see the world as it is, our mind processes the information
- Some mechanisms occur in a “pre-conscious” way
- Some actions are learnt, some others not



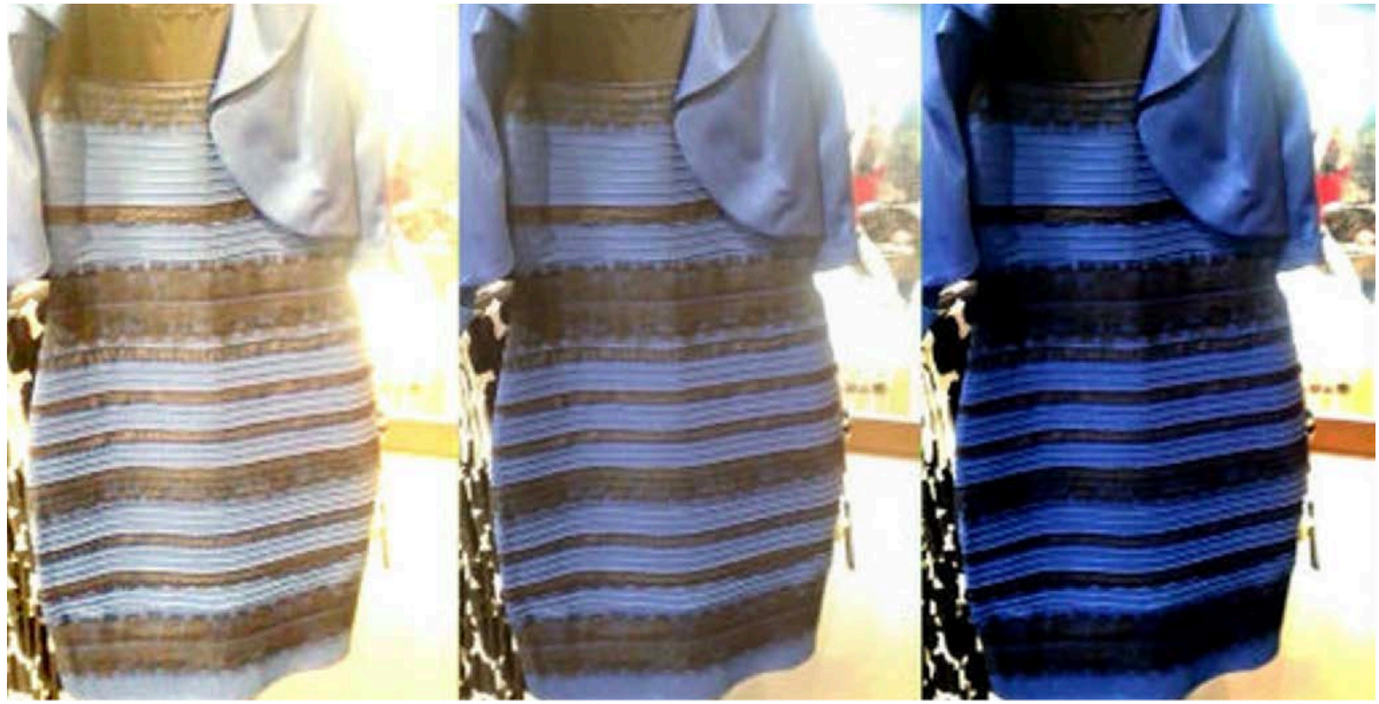

**Encuentra el punto rojo**



# Visual Perception

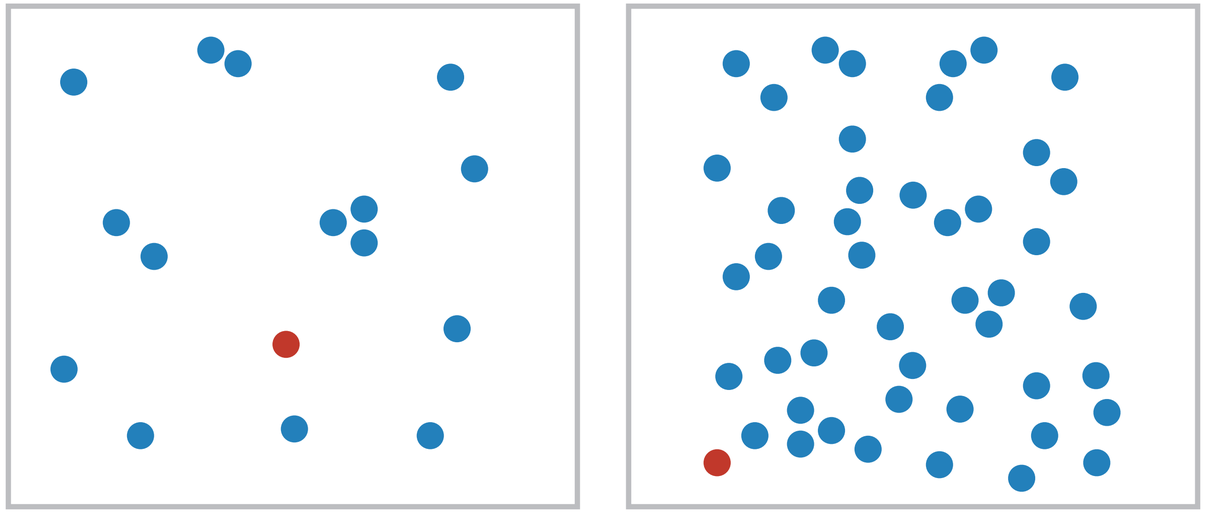
## The mind tricks

- Pop-out effect
- Gestalt principles
- Salience/Contrast

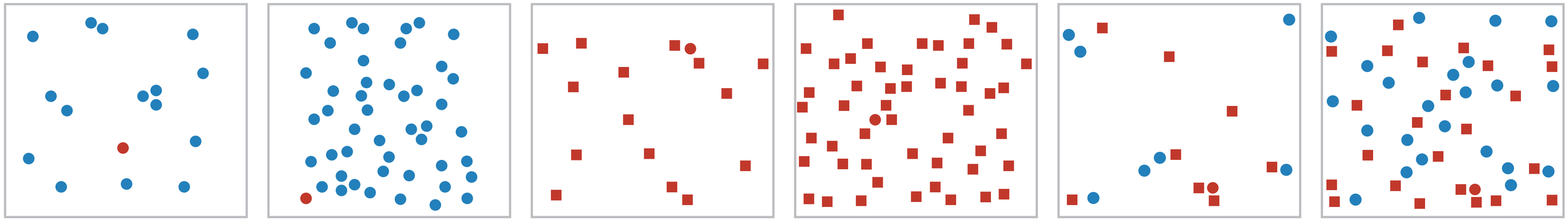


# Pop-out

Low level of visual perception = Early stage of processing = Very fast processing



# Pop-out

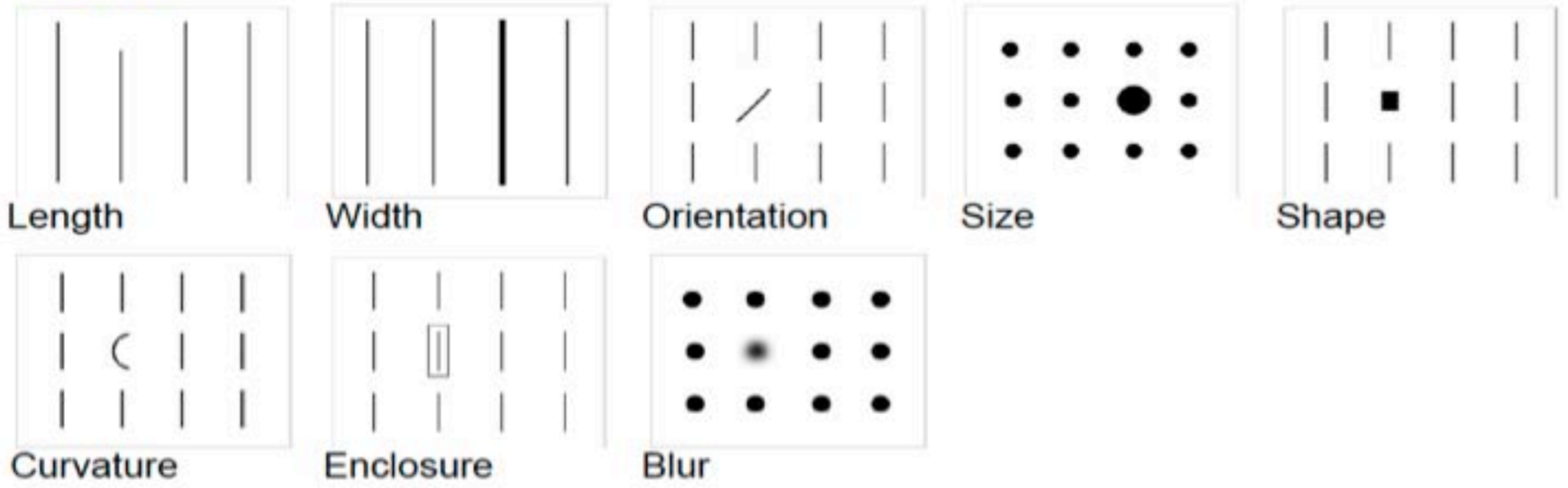


Speed detection in some channels is independent of distractor count
In other channels, speed depends on the amount of distractors

In mixed channels speed depends on both channel and amount of distractors.
Primary visual cortex can be tuned for circles or red things, not both

# Pop-out

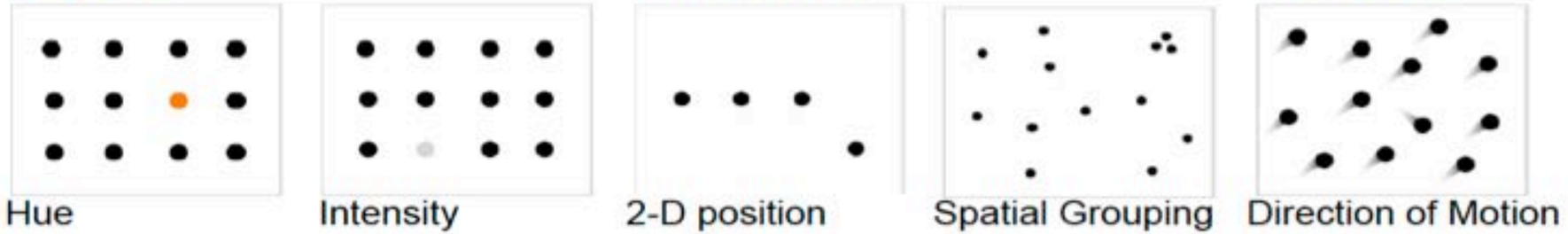
## Form




## Color

## Position

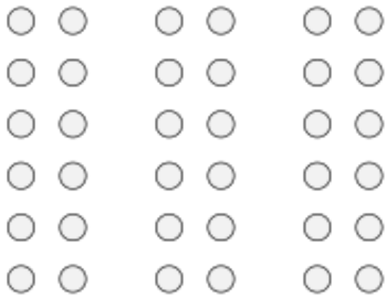
## Motion



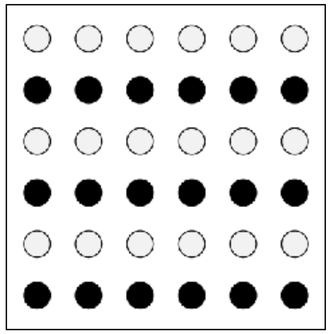

Colin Ware

# Gestalt principles

**Proximity**



**Similarity**



**Continuity**

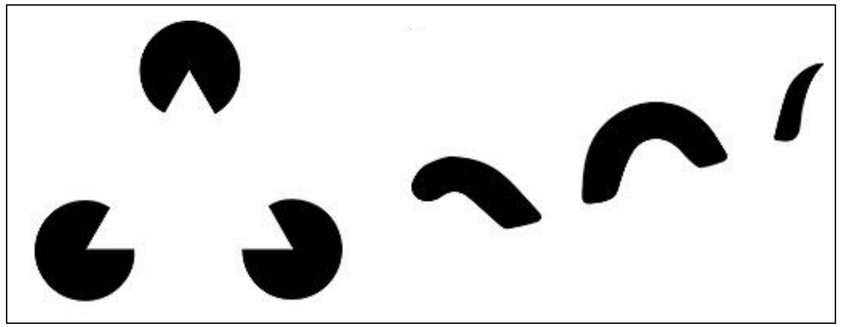
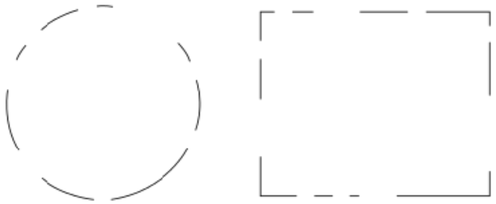
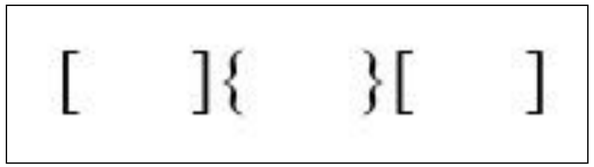


**Figure and ground**

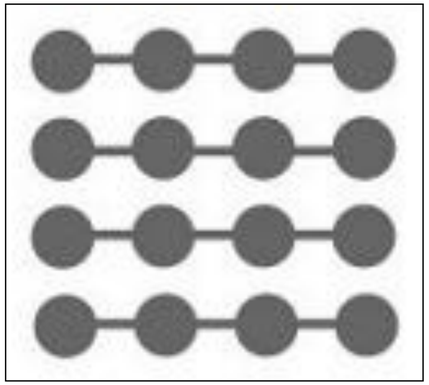
**Closure**



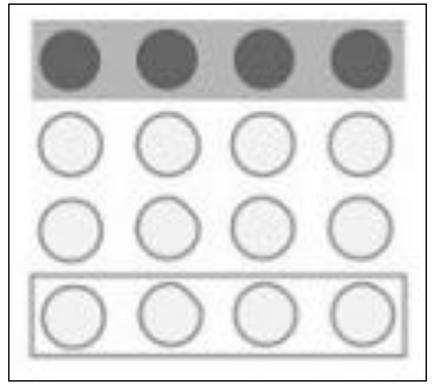
**Simmetry**



**Connection**

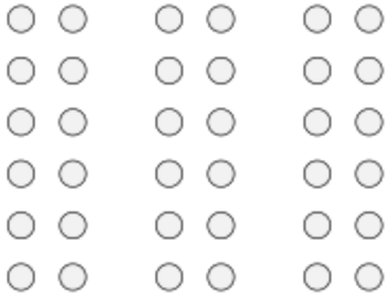


**Enclosure**

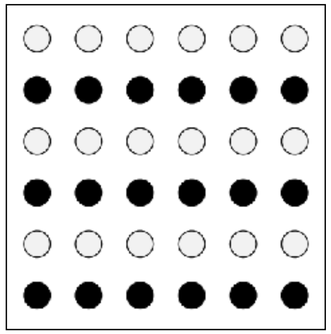


# Gestalt principles

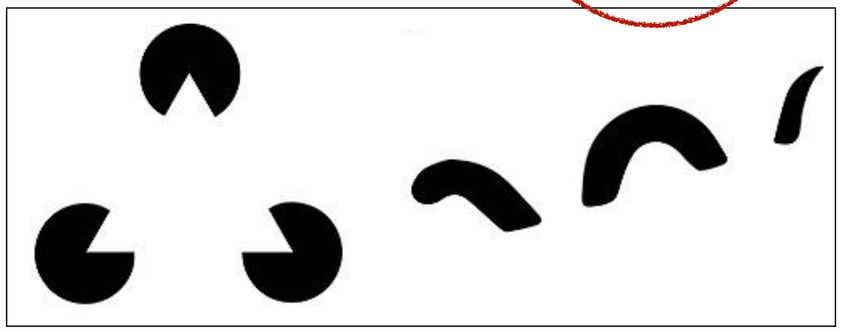
## Proximity



## Similarity

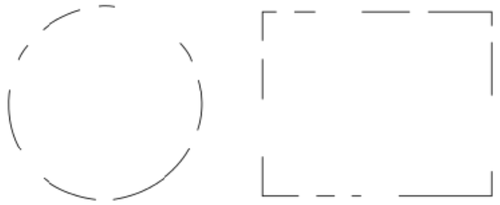


## Continuity

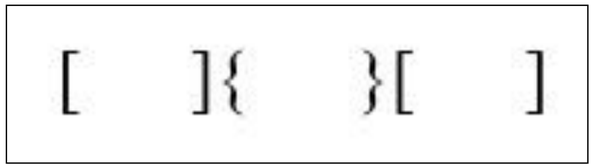


## Figure and ground

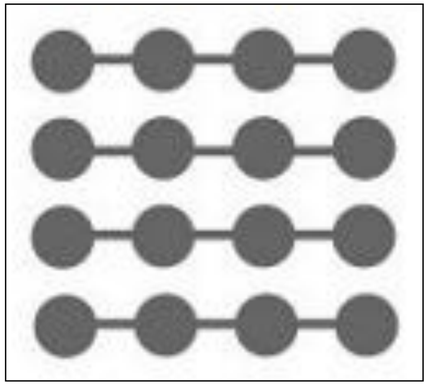
## Closure



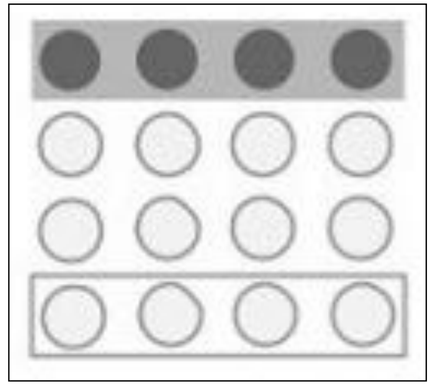
## Simmetry



## Connection



## Enclosure



# Gestalt principles

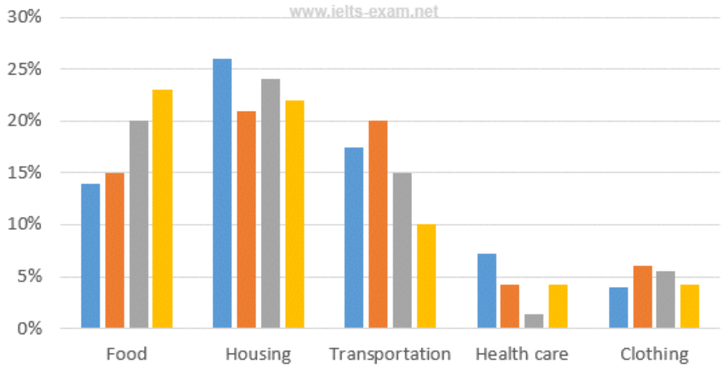
## Proximity

1234  56789 1234

1234  56789 1234
123 2567 8912341

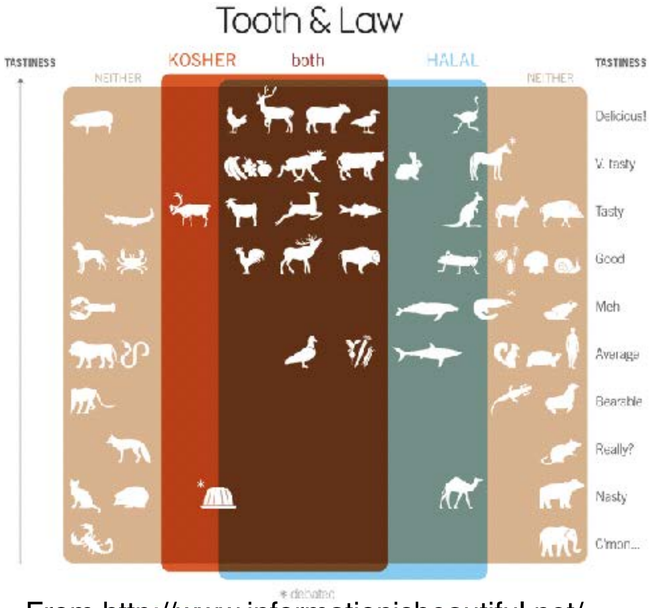
1234 6789 12 234

## Similarity



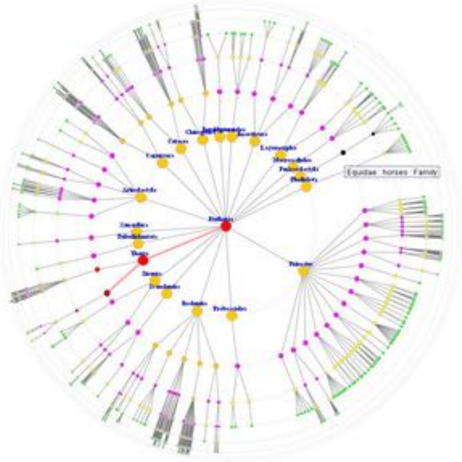

## Connectedness

## Closure




From http://www.informationisbeautiful.net/

## Continuity



# Grouping

Visual attributes which serve to arrange elements, and create categories and relations

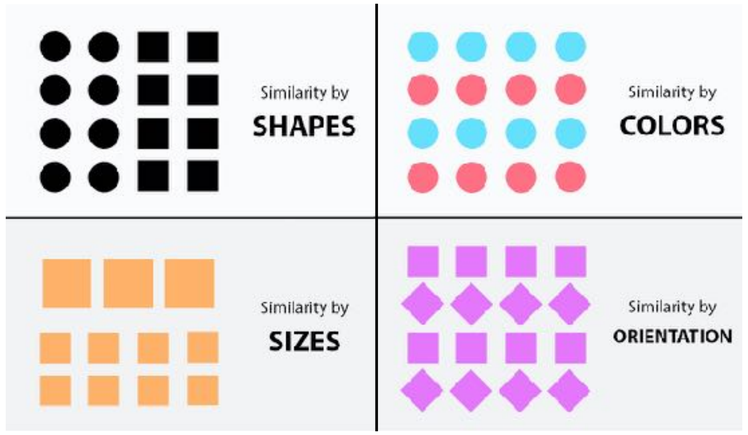
## Proximity

1234  56789 1234

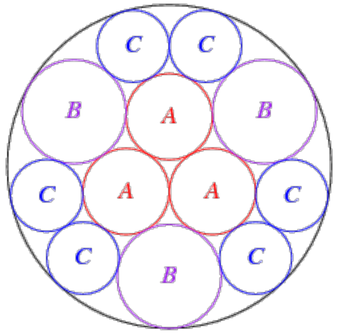
1234  56789 1234
123 2567 8912341

1234 6789 12 234

## Similarity




## Containment



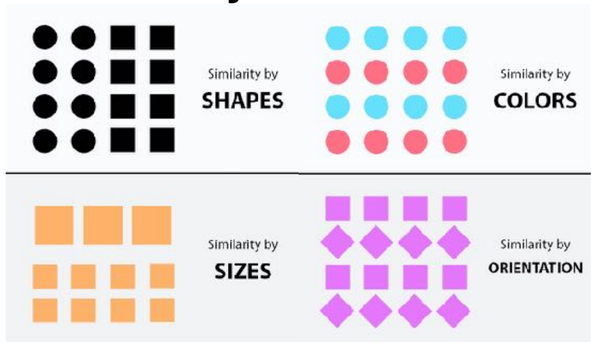
## Proximity

1234 56789 1234

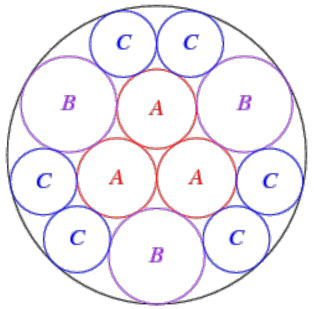
1234 56789 1234
123 2567 8912341

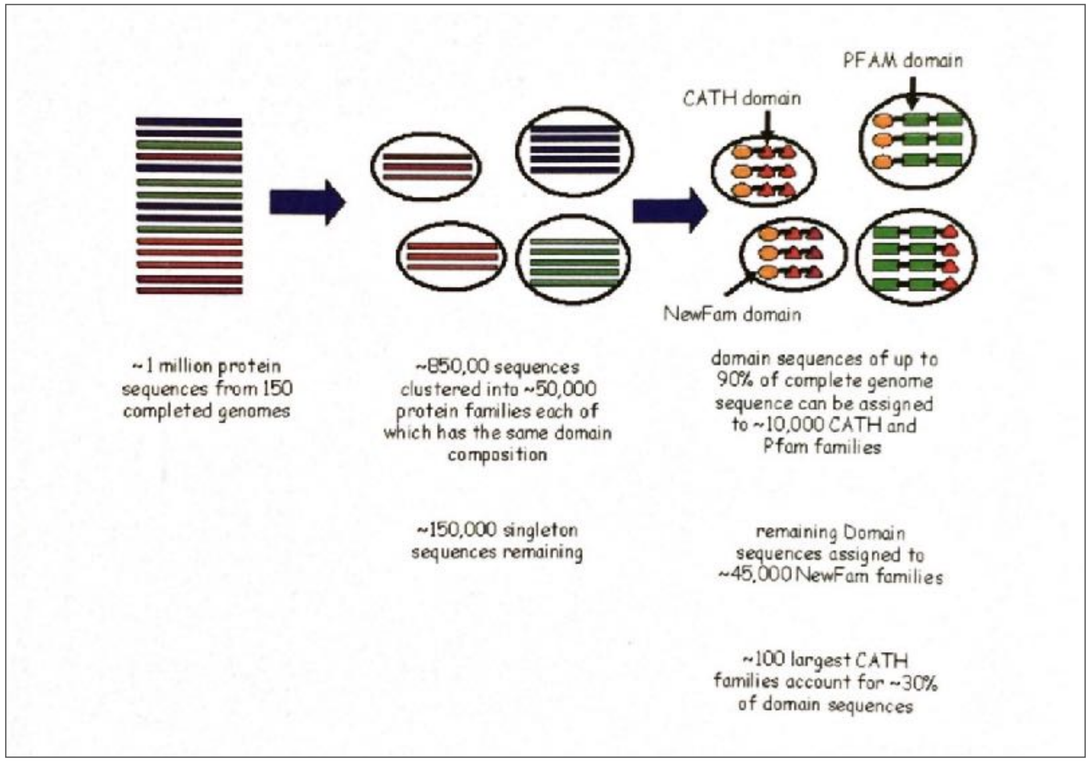
1234 6789 12 234

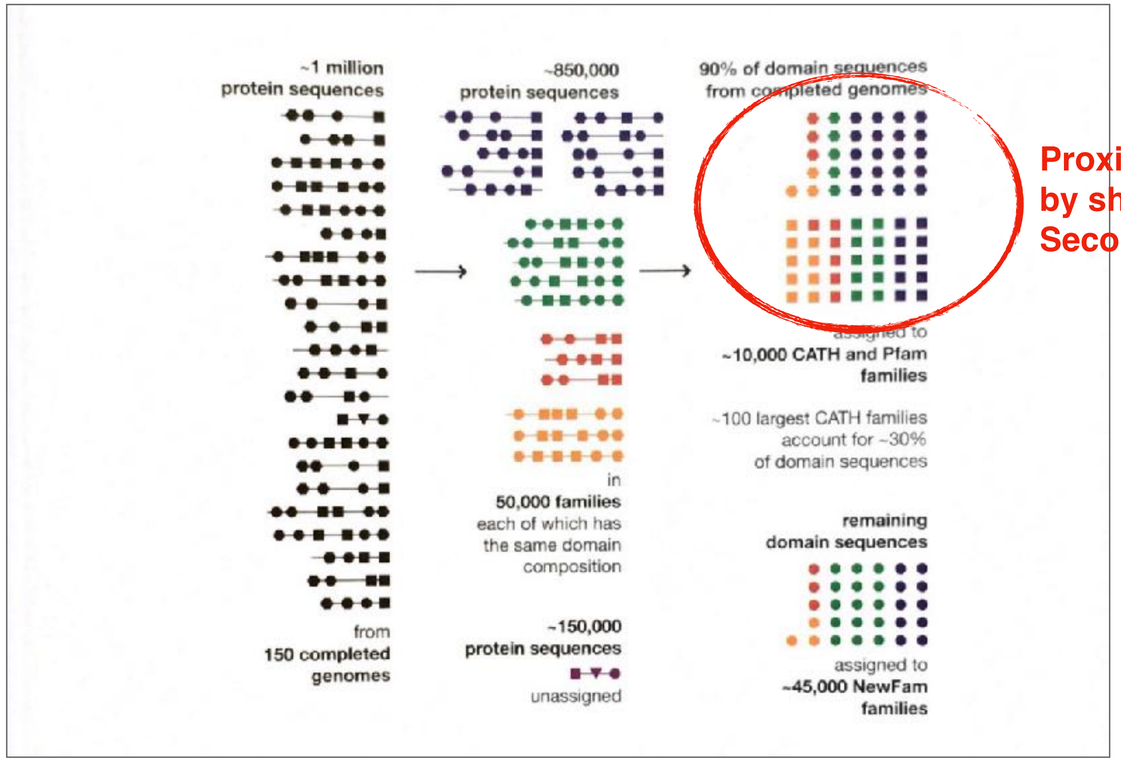
## Similarity




## Containment



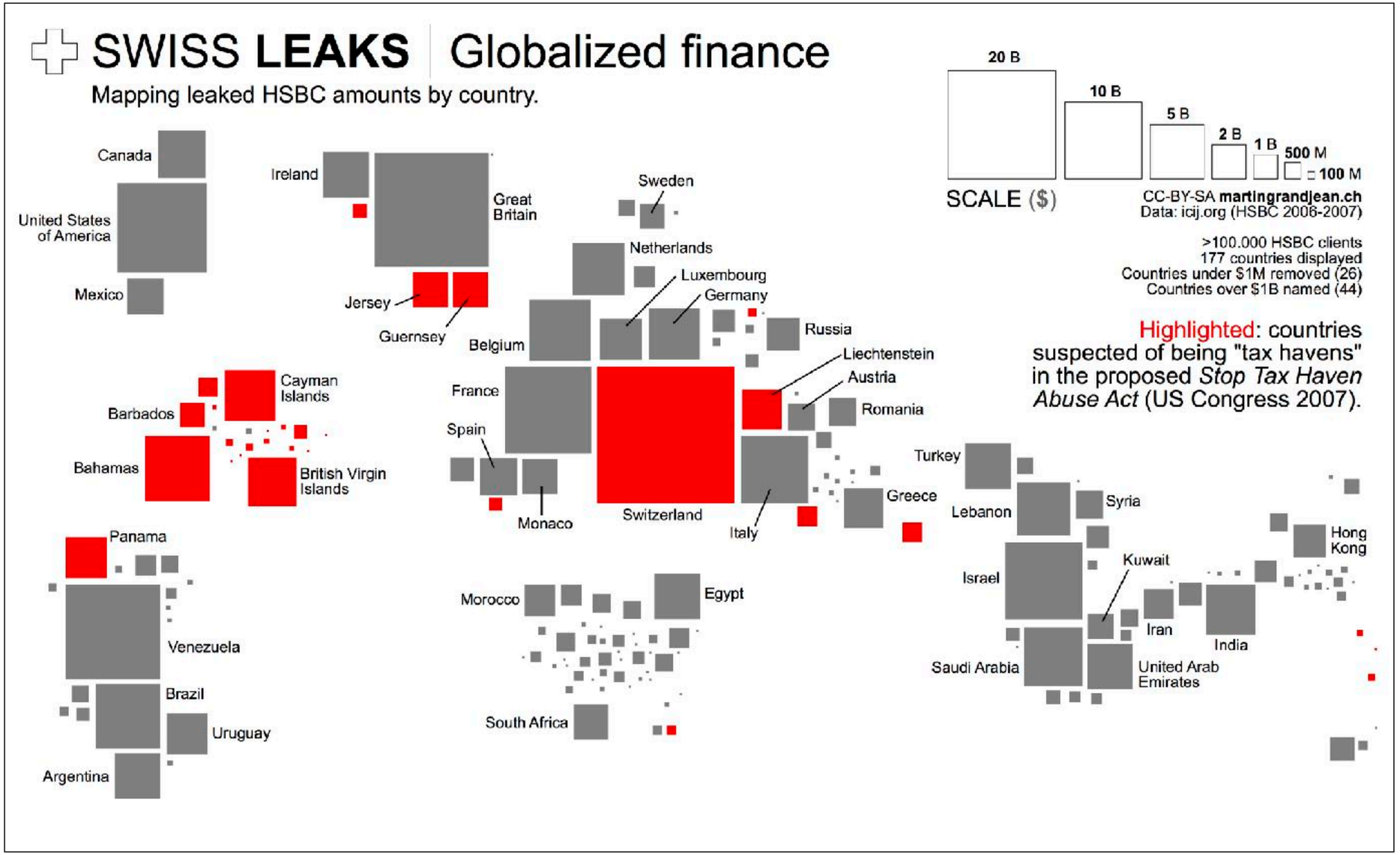

**Proximity reinforced by shape similarity**
**Second level by color**

# Contrast is as important as other visual attributes

**Easy**

**Difficult**

# If something has to be easy to find, make it different!



**Channels:** Expressiveness Types And Effectiveness Ranks

## Magnitude Channels: Ordered Attributes

Position on common scale

Position on unaligned scale

Length (1D size)

Tilt/angle

Area (2D size)

Depth (3D position)

Color luminance

Color saturation

Curvature

Volume (3D size)

Same

Same

Effectiveness

Best

Least

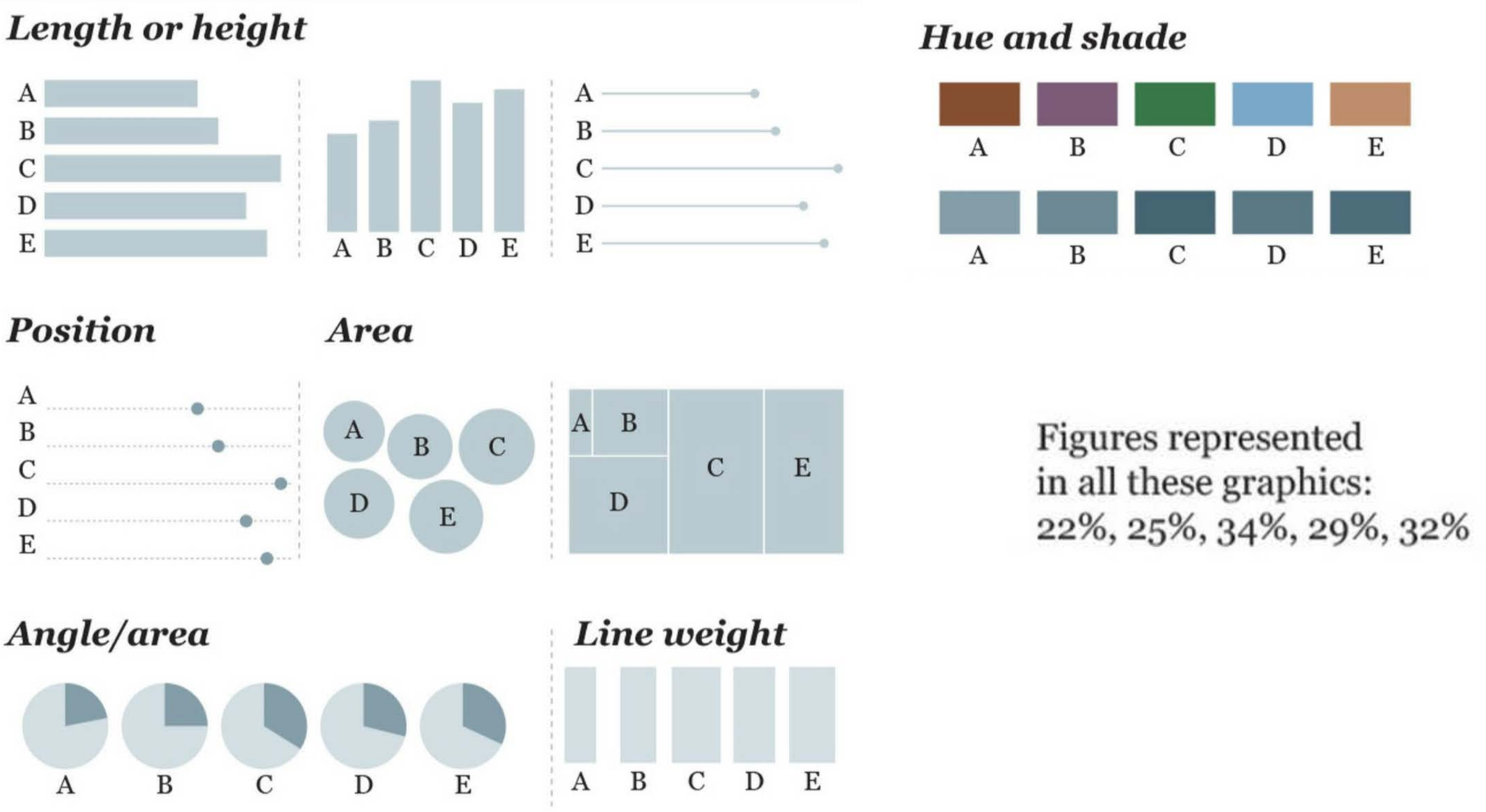
## Identity Channels: Categorical Attributes

Spatial region

Color hue

Motion

Shape

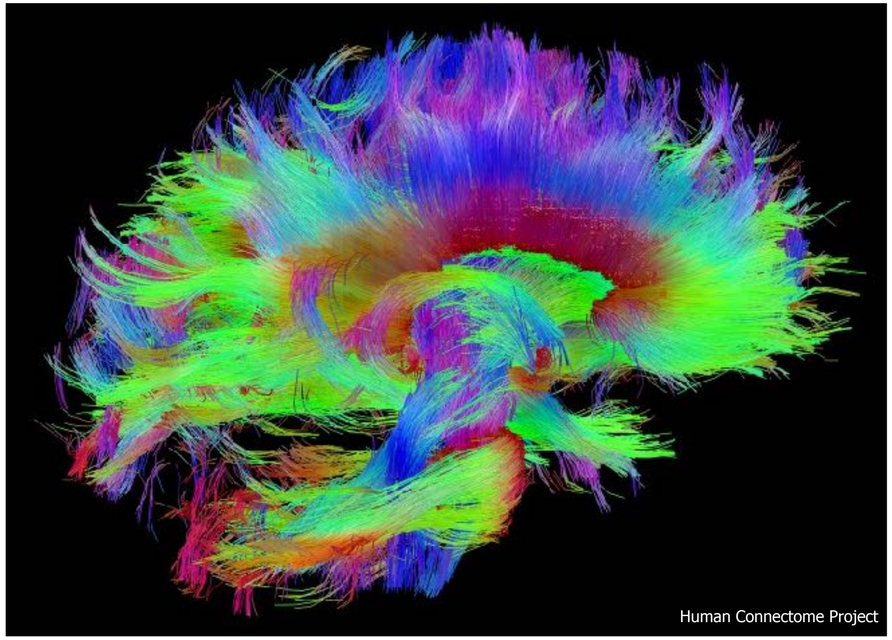


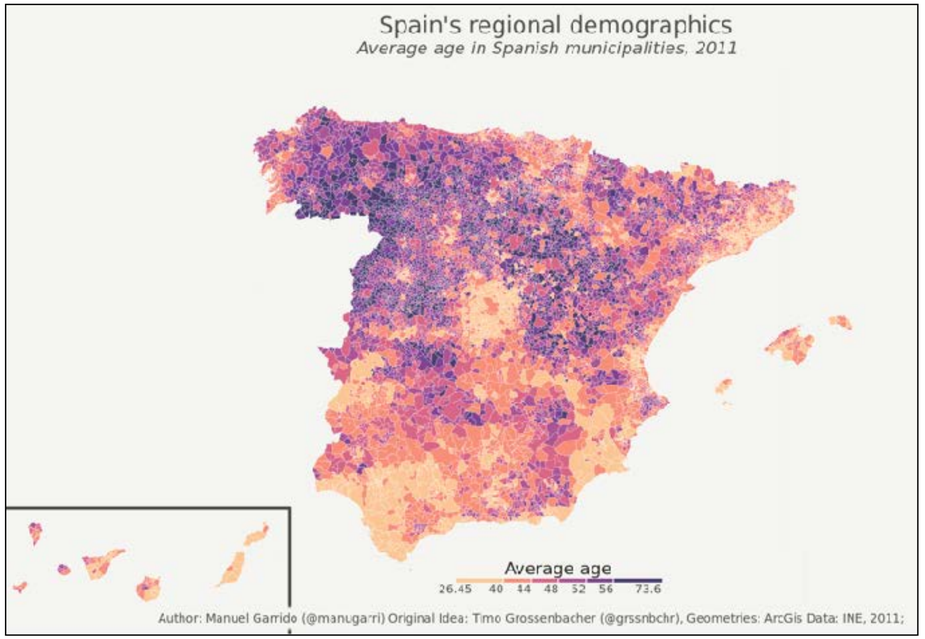

from The Truthful Art, Cairo, A.

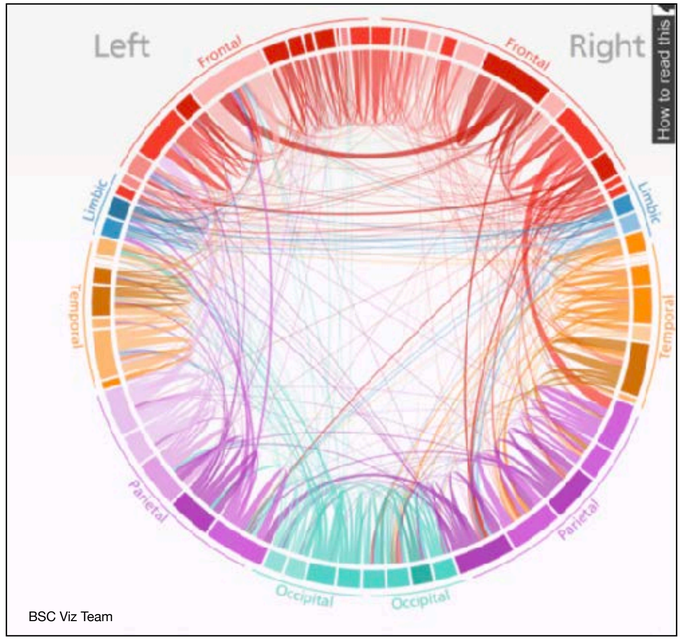
# Graphic design: **Color**

# Color scales

## Match color to data

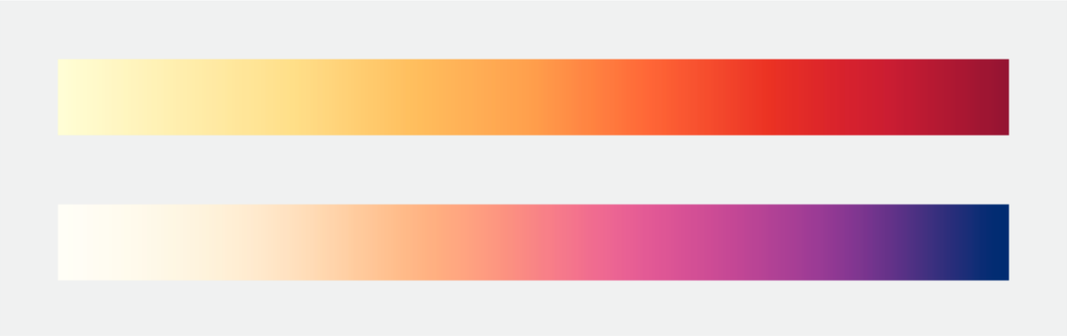


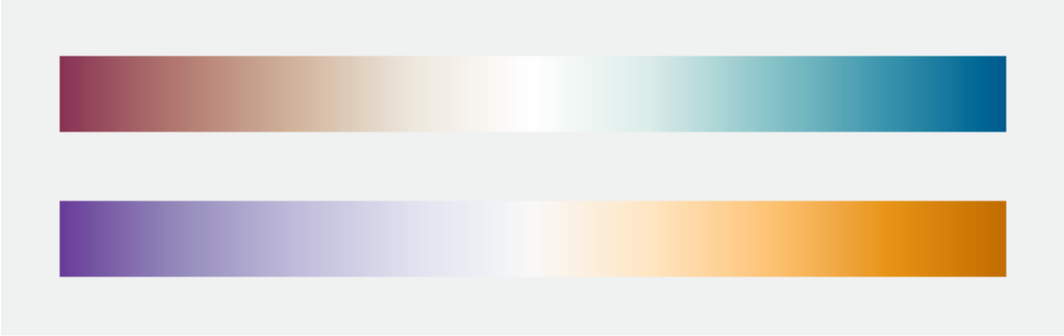



# Match color to data

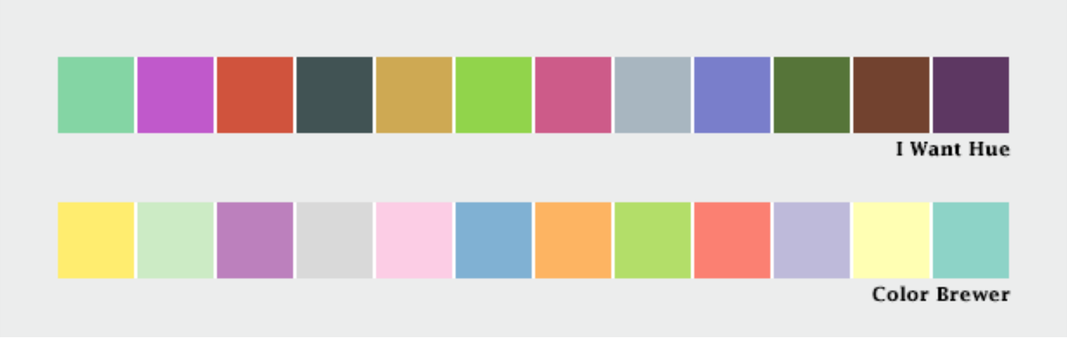
## Sequential



## Divergent

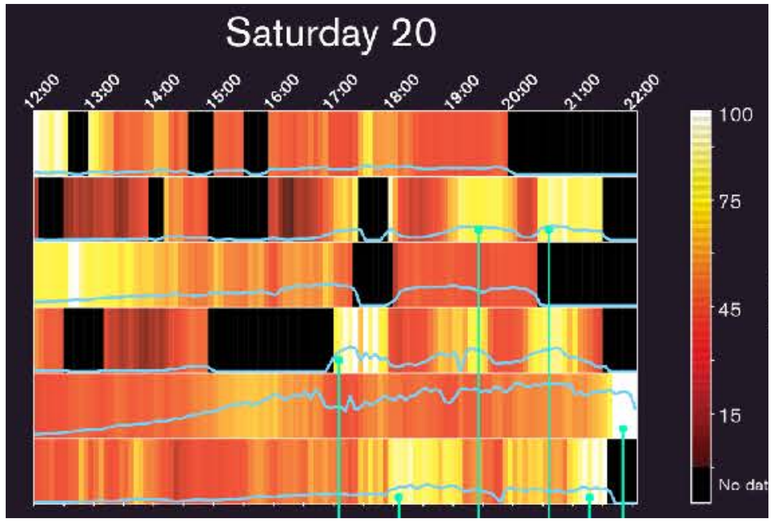


## Categorical



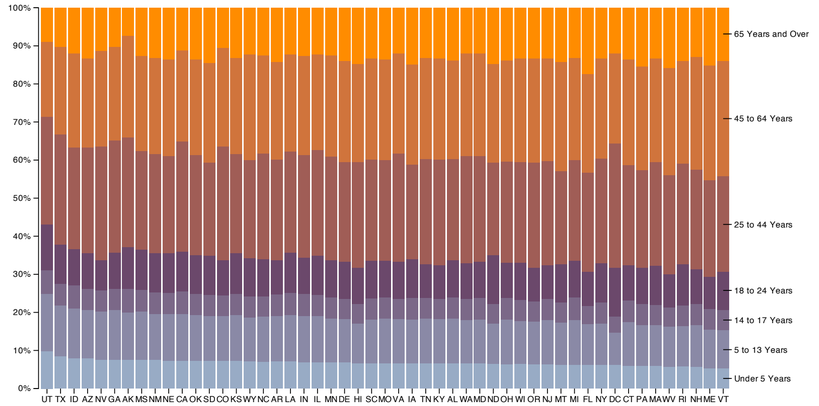
# Sequential

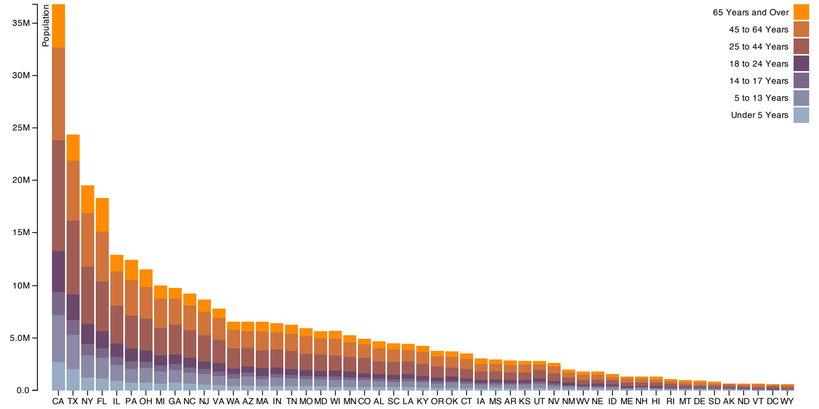
## Discrete



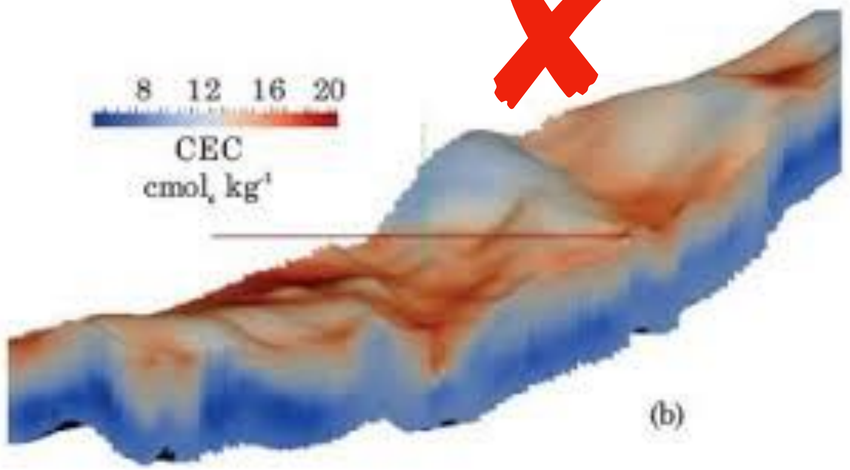

## Continuous

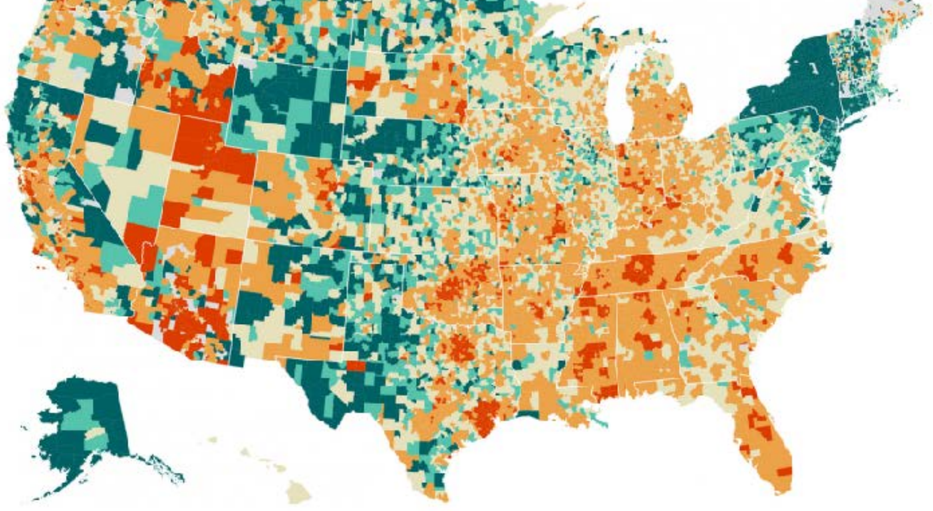
## Ordinal

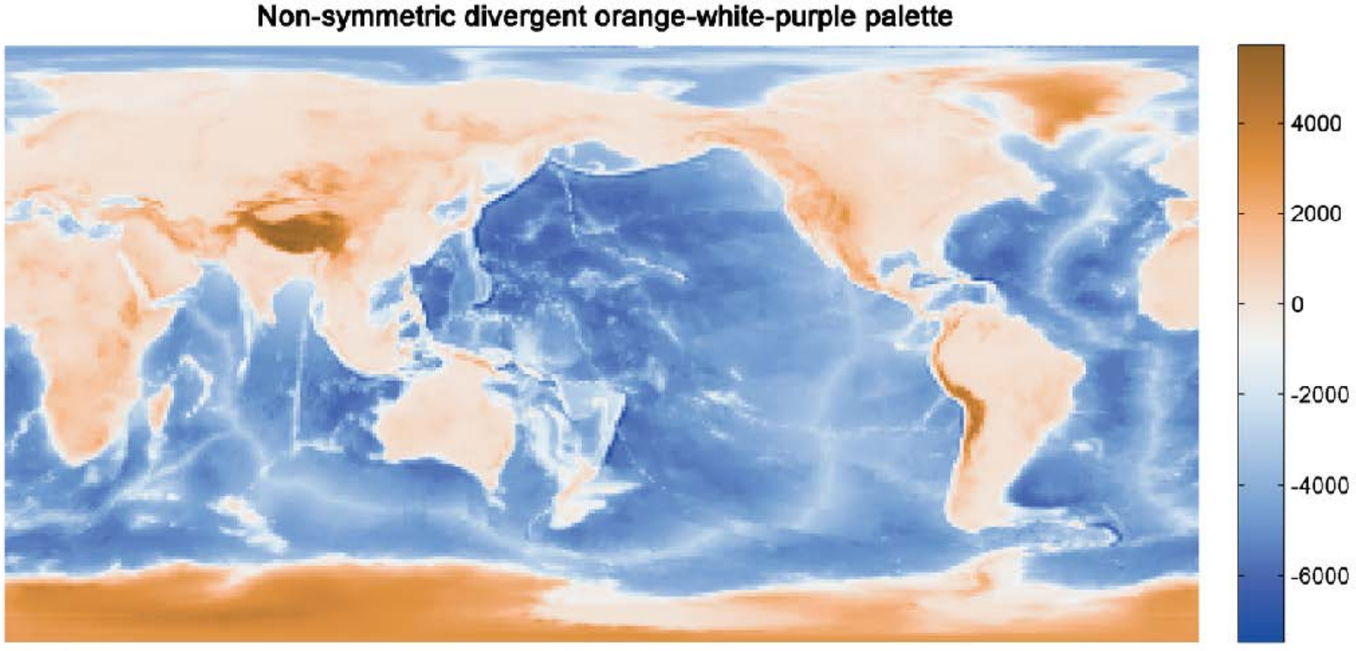


# Divergent

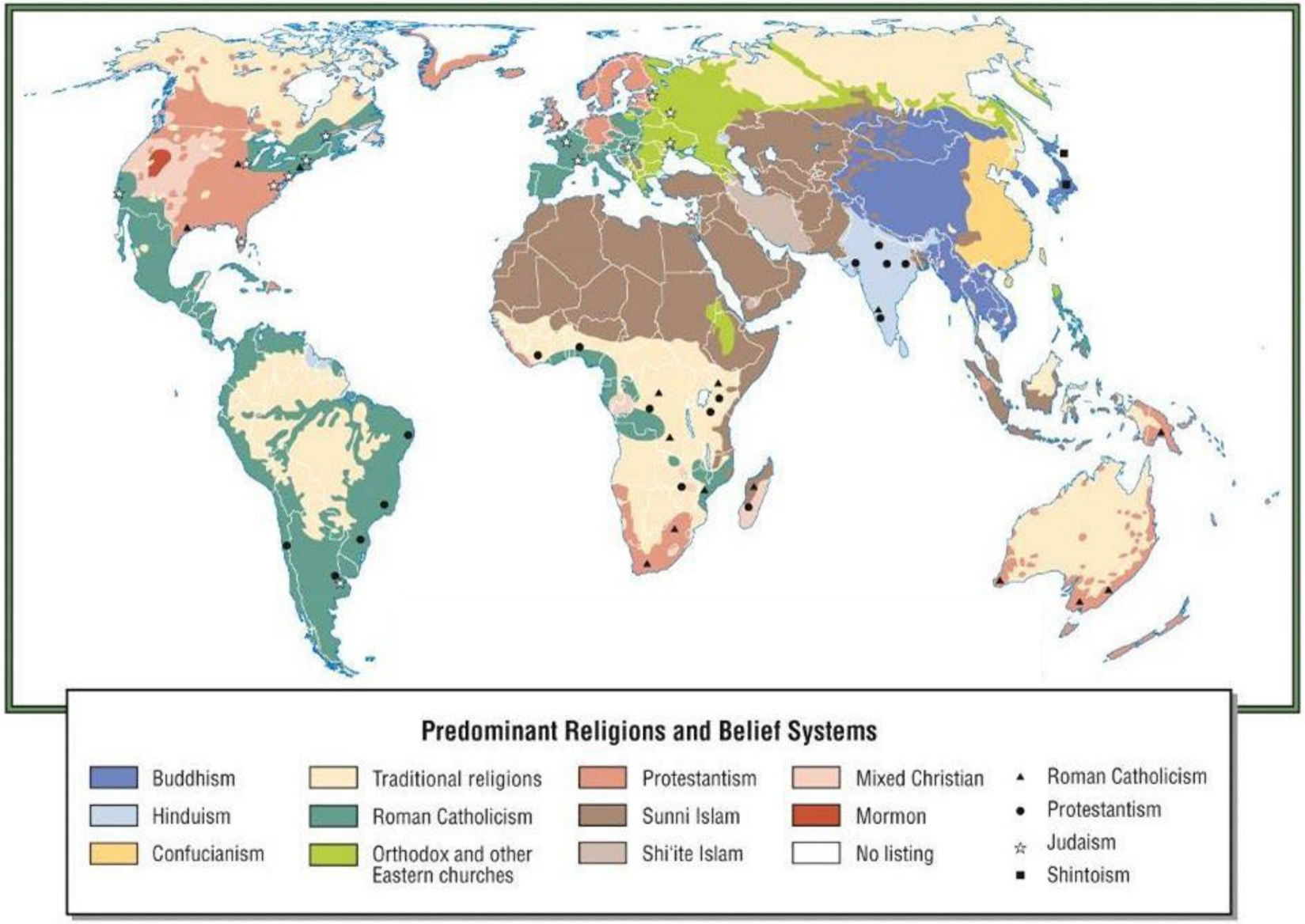

8 12 16 20
CEC
cmol$_c$ kg$^{-1}$
(b)


SPENDING PER STUDENT, BY SCHOOL DISTRICT
Adjusted for regional differences, for primary and unified school districts
National average: $11,841
-33%
-10%
+10%
+33% of national average


Non-symmetric divergent orange-white-purple palette
4000
2000
0
-2000
-4000
-6000

# Categorical

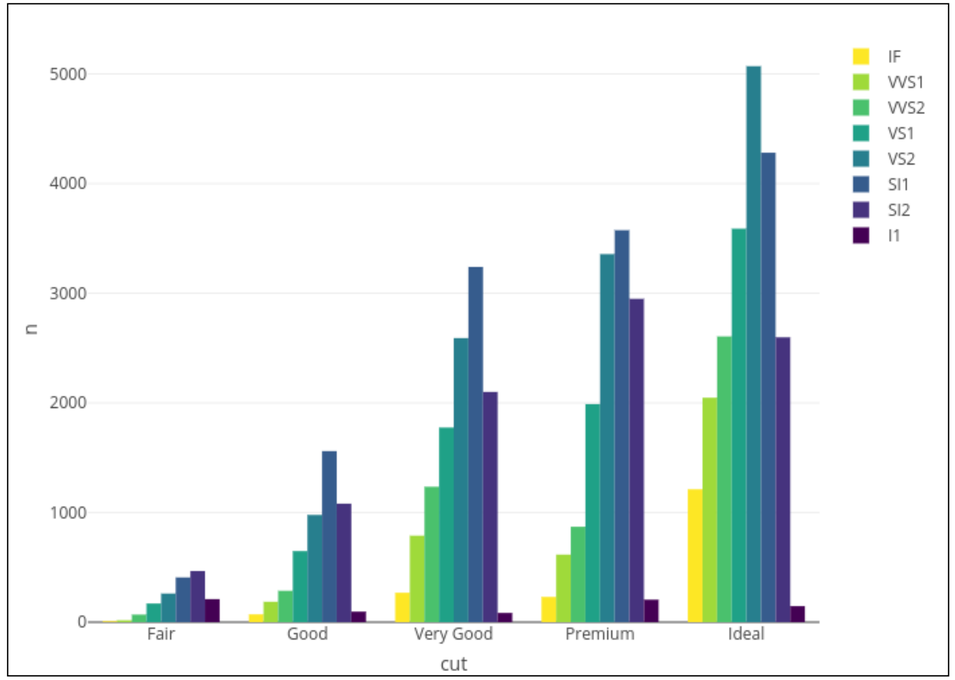
**No more than 12 colors**

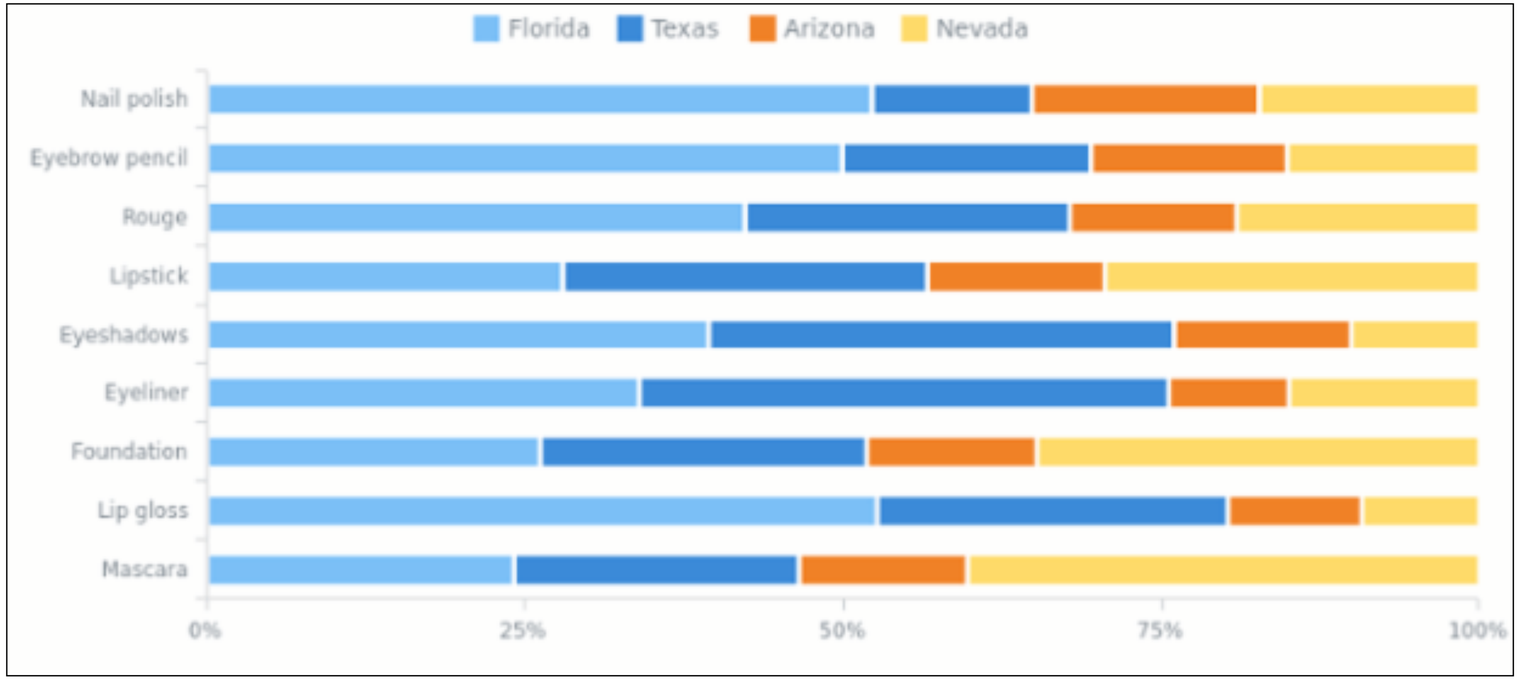


# Categorical

Beware of continuous gradients, can be confused with ordinal

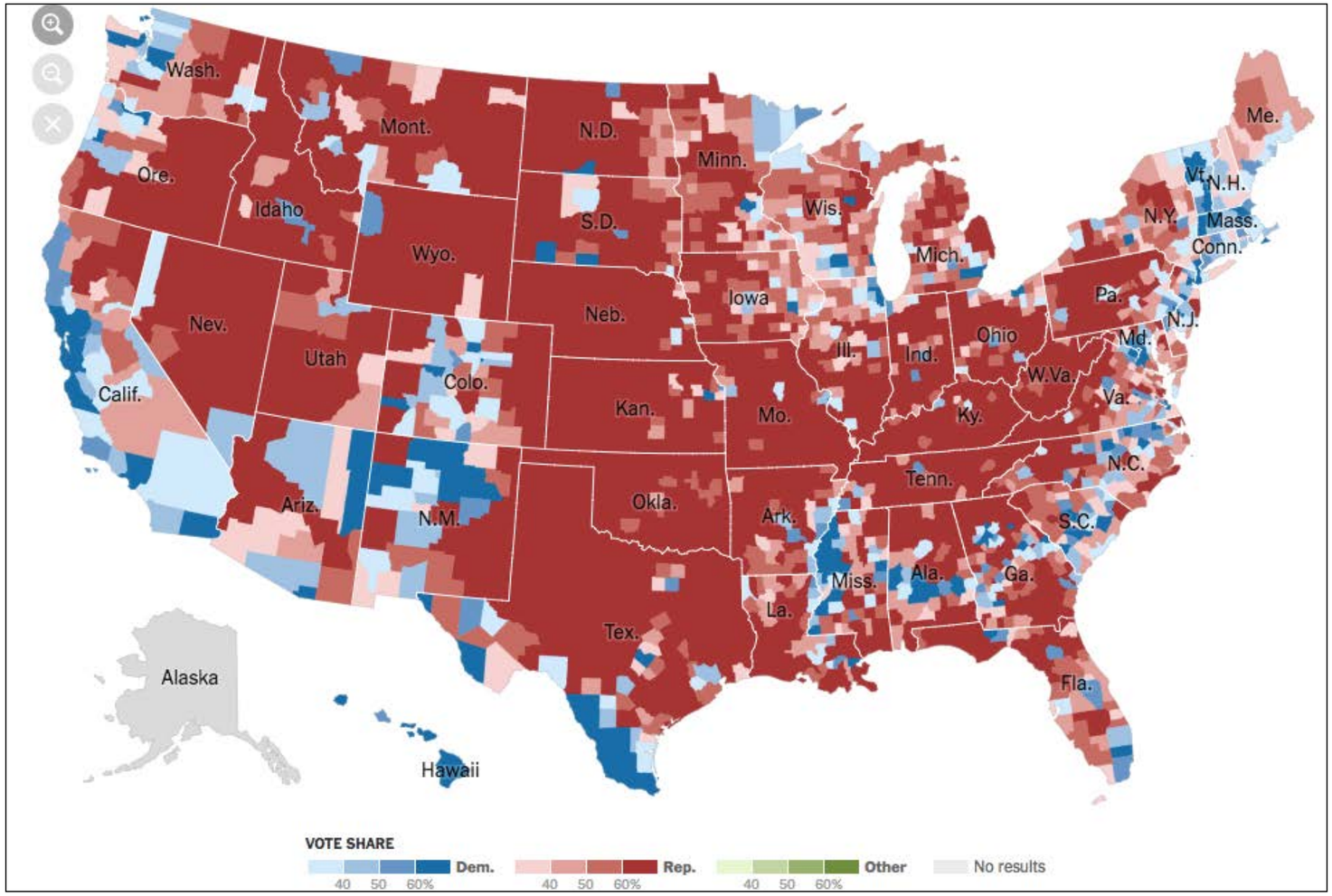
Better use big differences in hue, sat, and/or luminosity



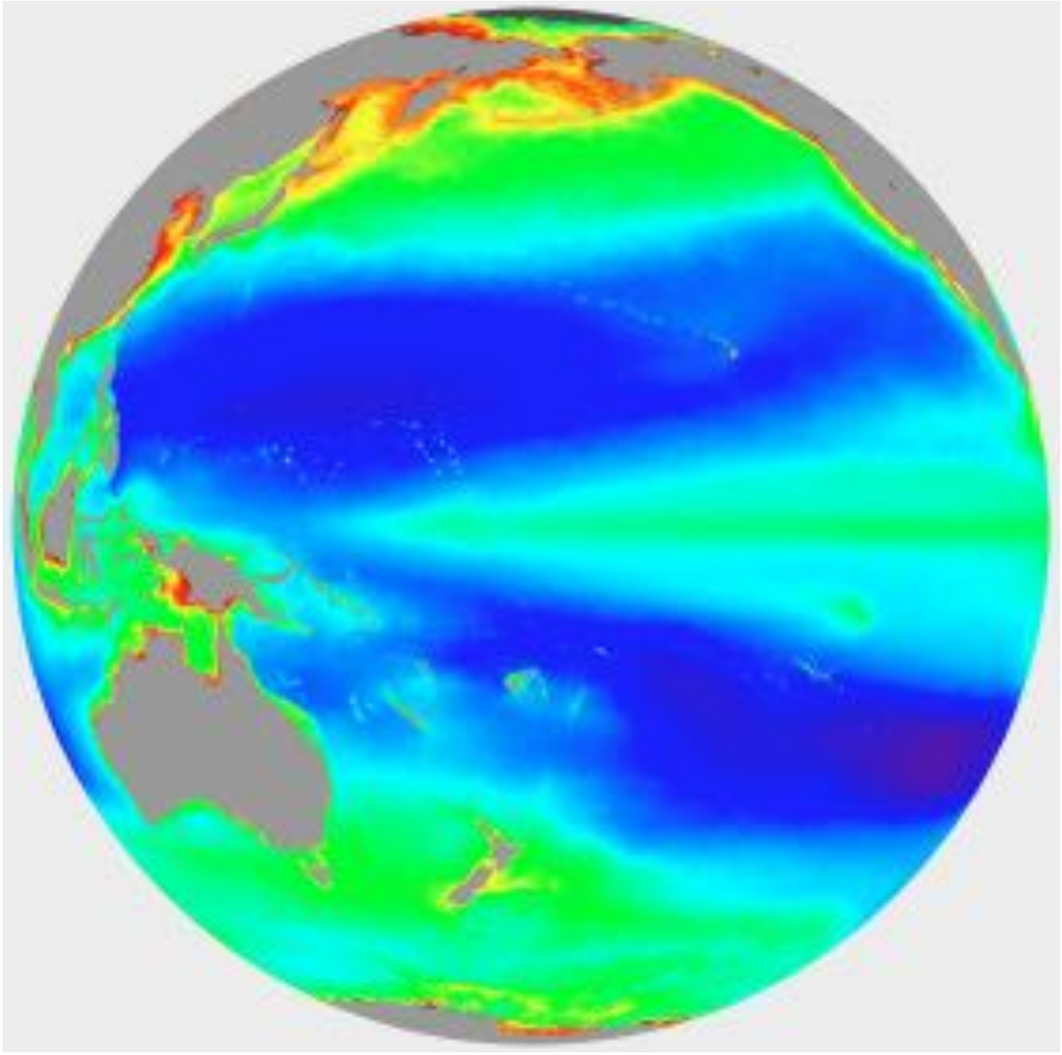
# Grouped

Grouped color scheme of 4 categories (4 hues) with 4 steps of saturation and lightness each




nytimes.com

# The rainbow color scale

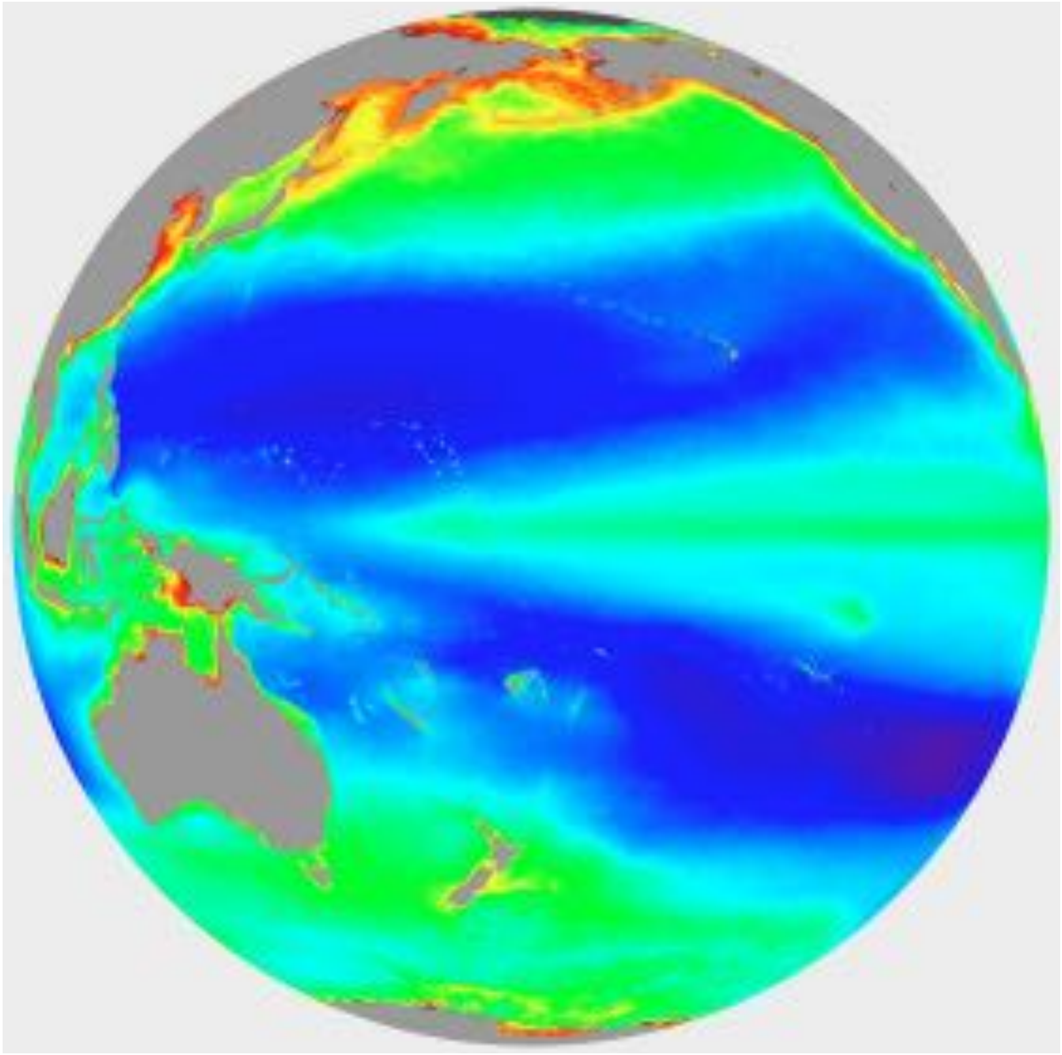


# The rainbow color scale

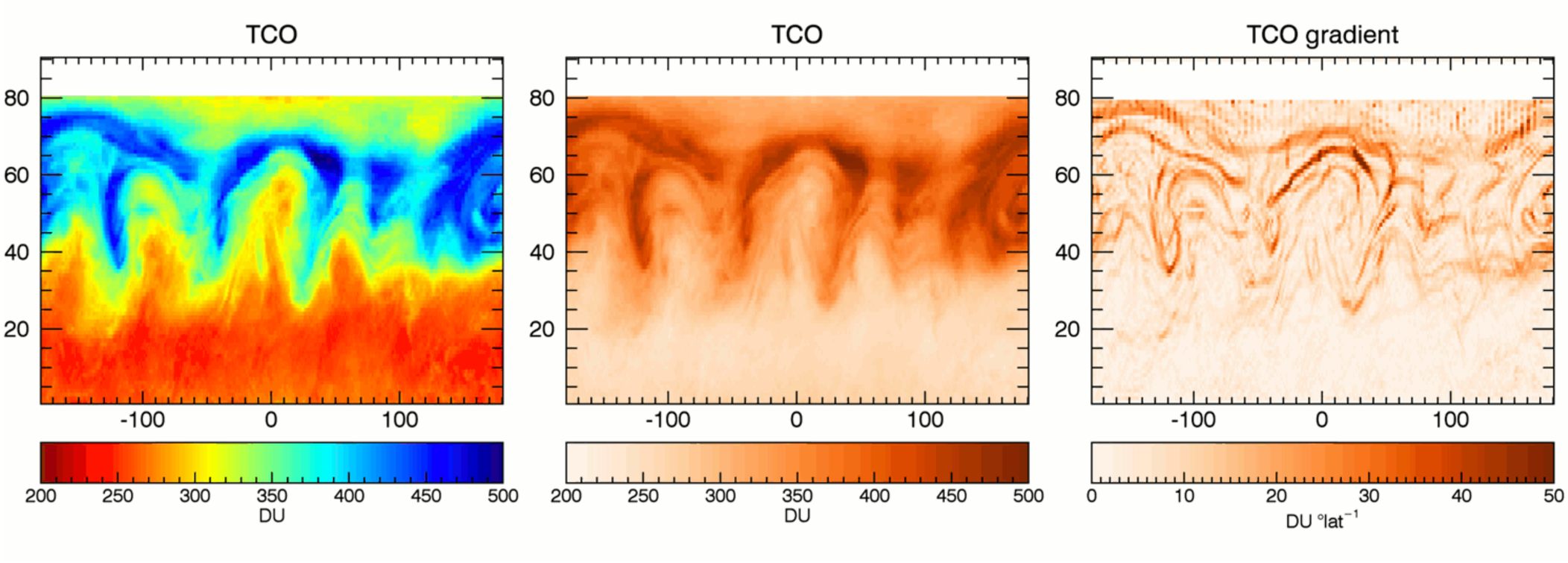
- We see some colours brighter than others
- Interpolations are not perceptually linear
- No continuous variation in lightness
- No inherent meaning

Not made for humans

**Don’t use it**



# The wrong color scale can show effects that are not present in the data

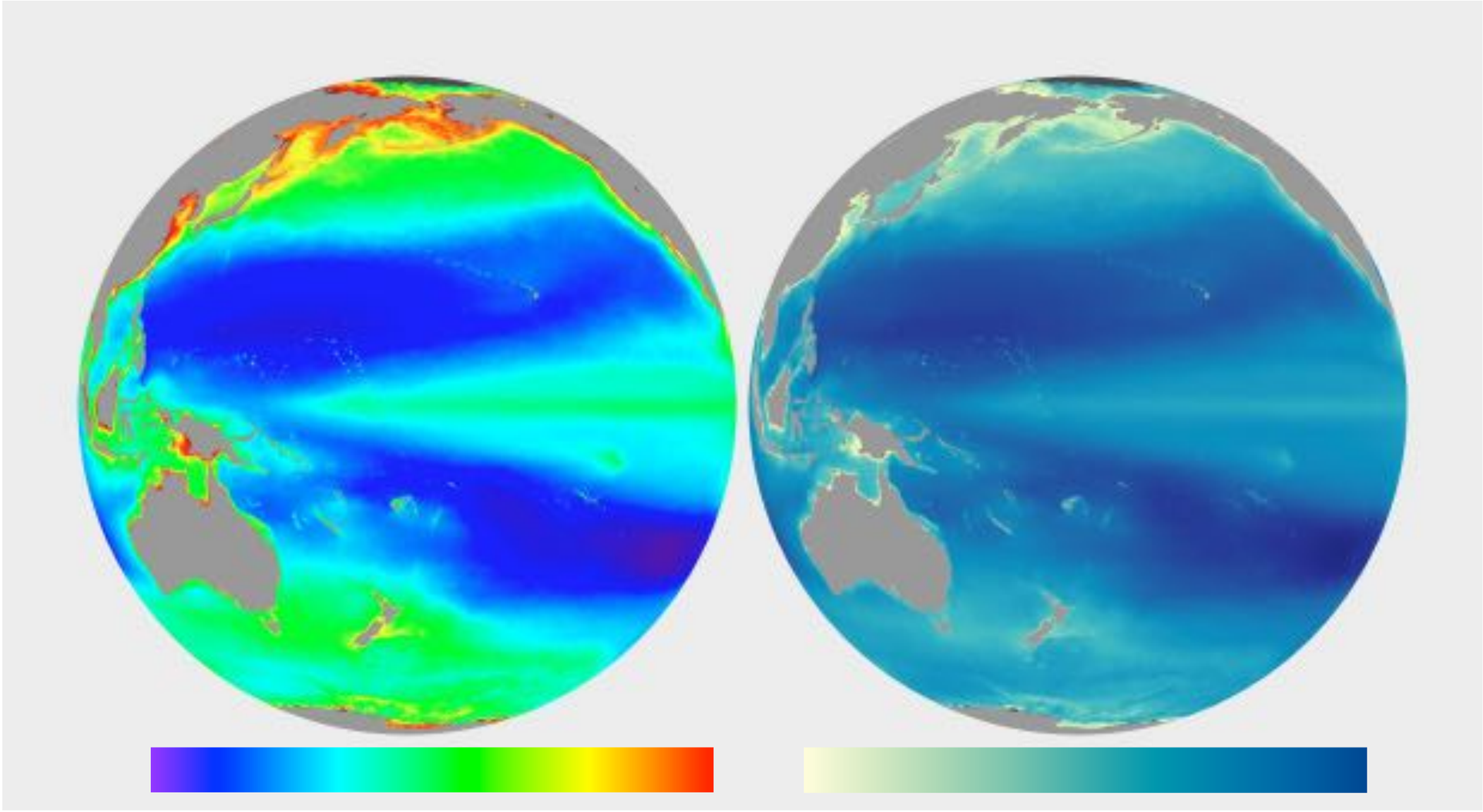



Sean Davis, NOAA

# Perceptually linear scales

Use scales with an even variation.

Scales that show contrast and change when there is change in the data and not because of optic distortions

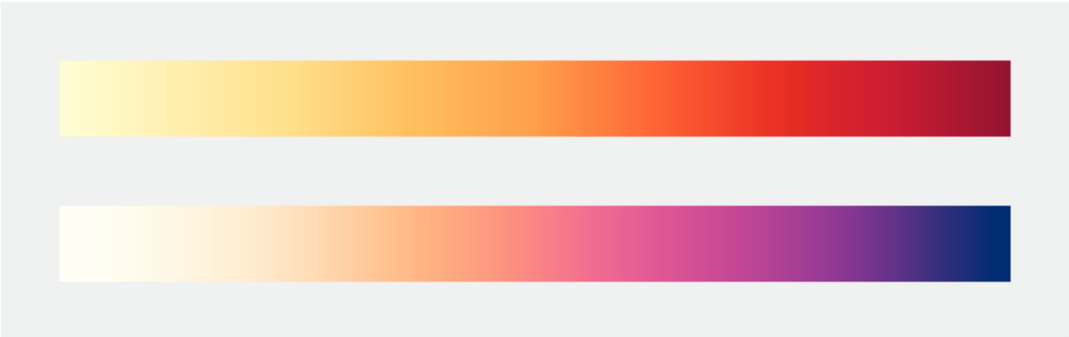


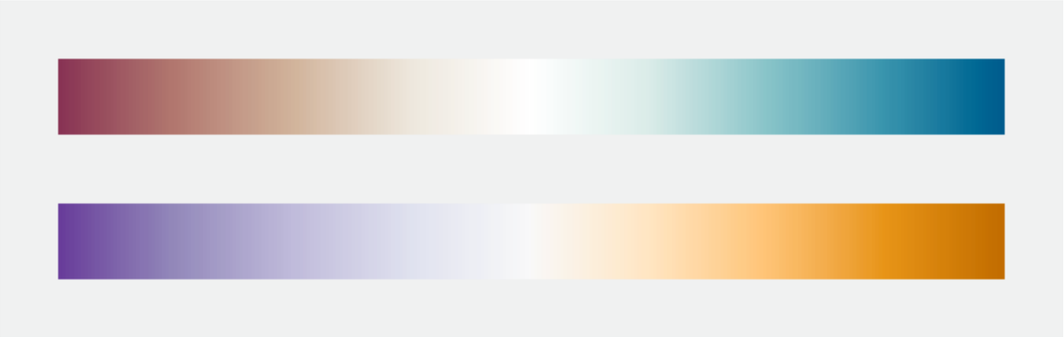
Simmon, 2013

# Perceptually linear scales

A good color scale should vary consistently across the range of values

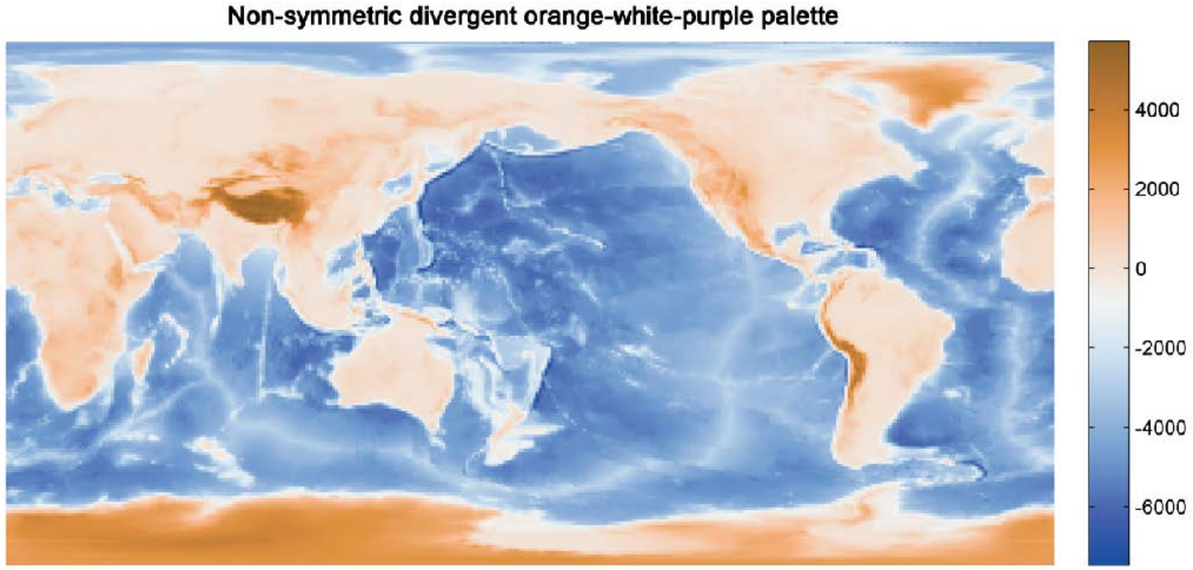
https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html



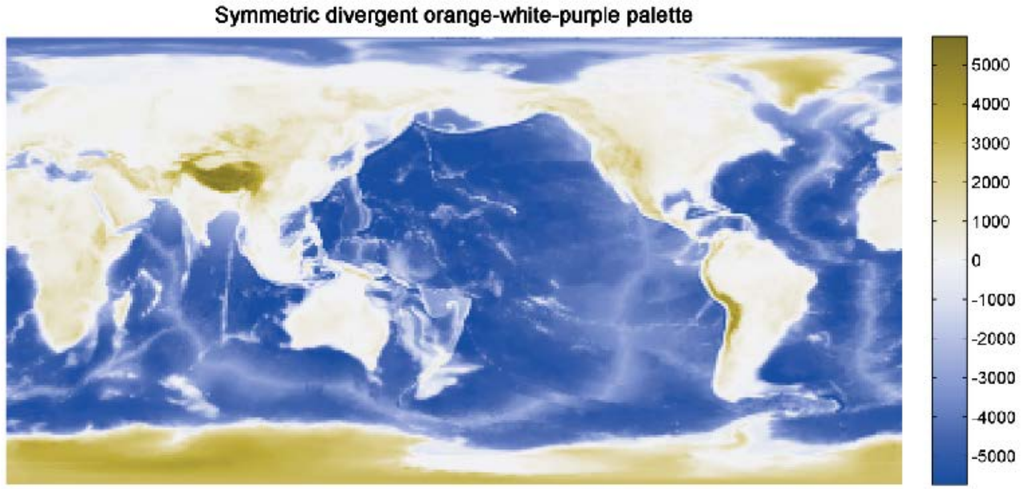


# Color deficiency

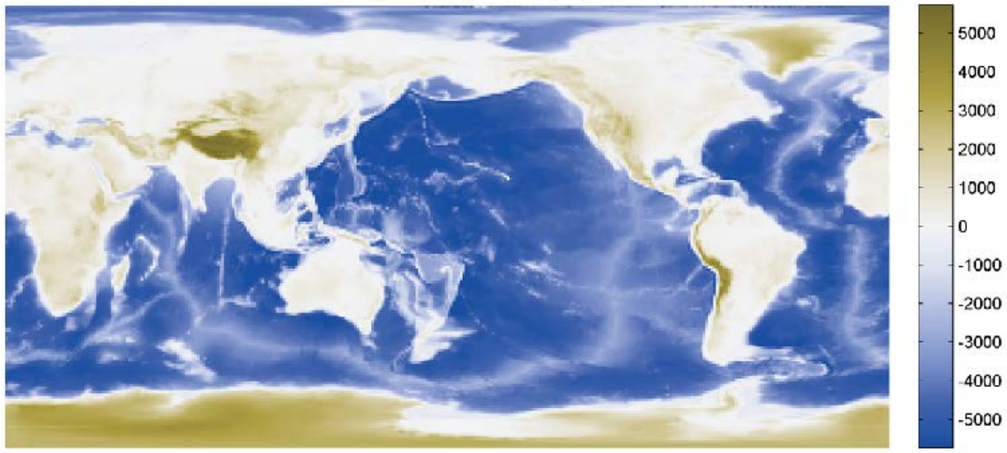
www.color-blindness.com/coblis-color-blindness-simulator/



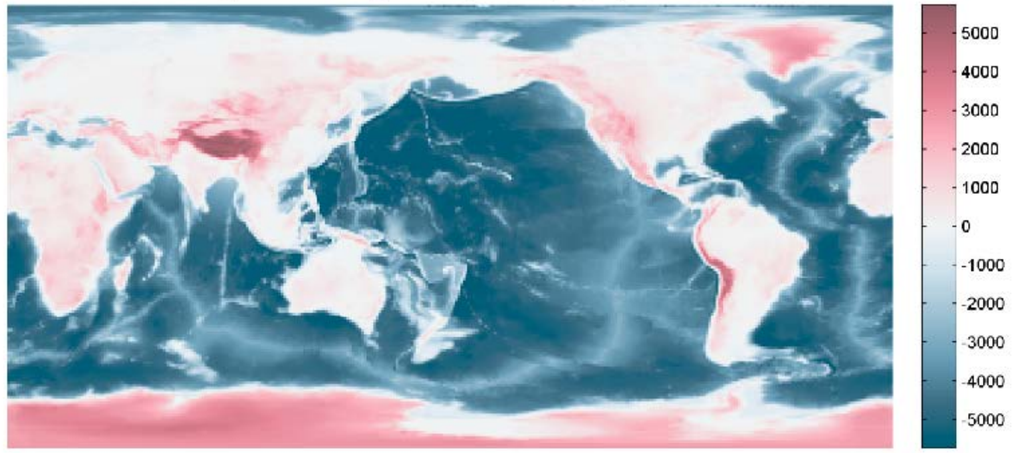

## Deuteranope




## Protanope

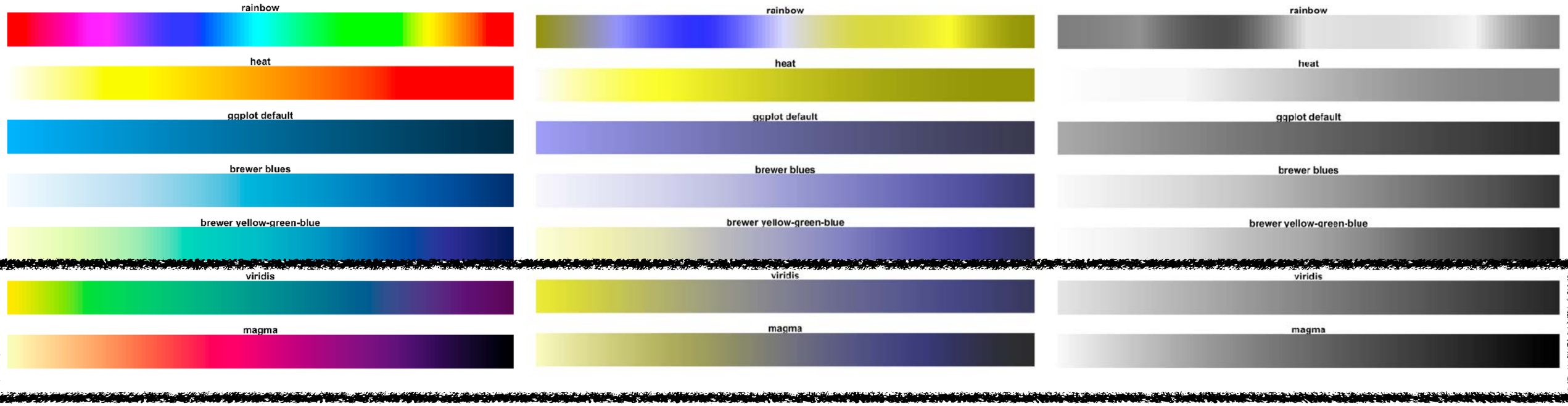



## Tritanope



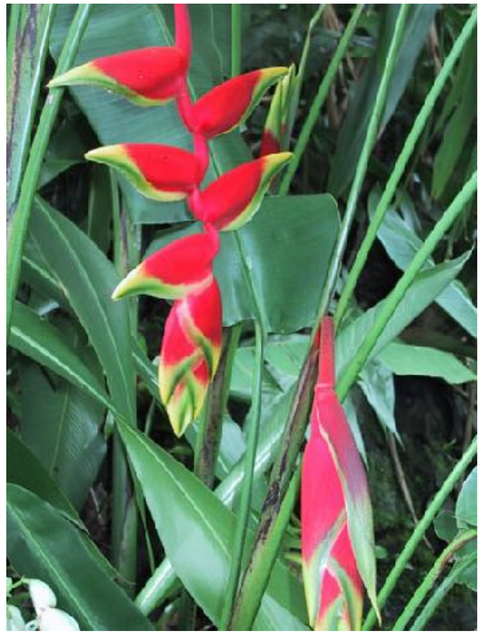
# Color scales

https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html

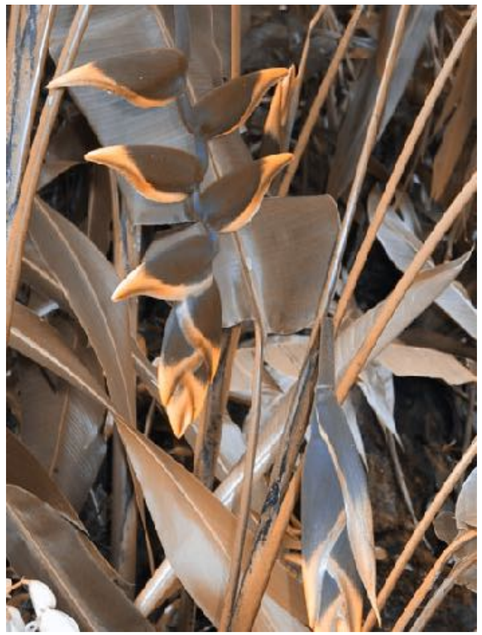


# Designing for color deficiency: Check with simulator

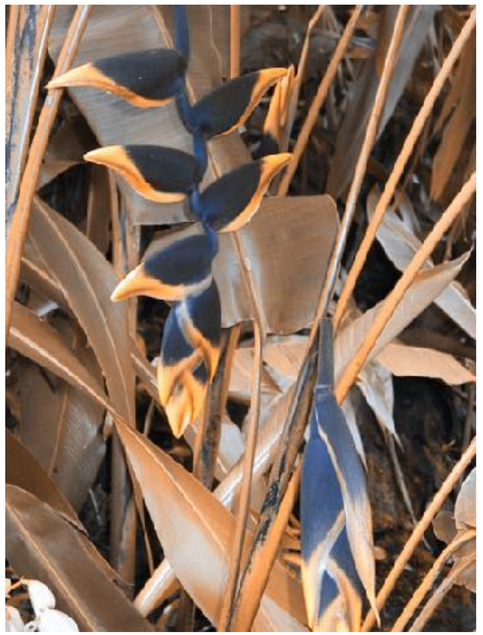
www.color-blindness.com/coblis-color-blindness-simulator/



Normal vision



Deuteranope

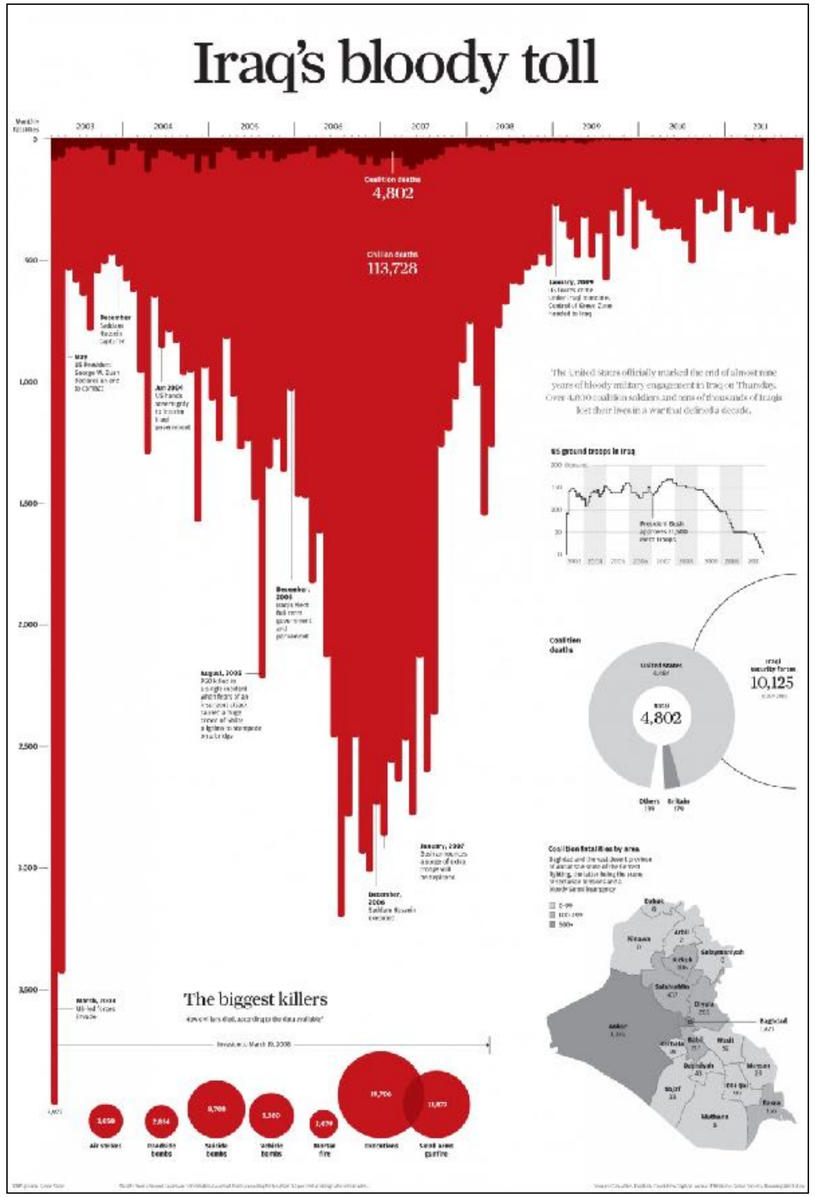


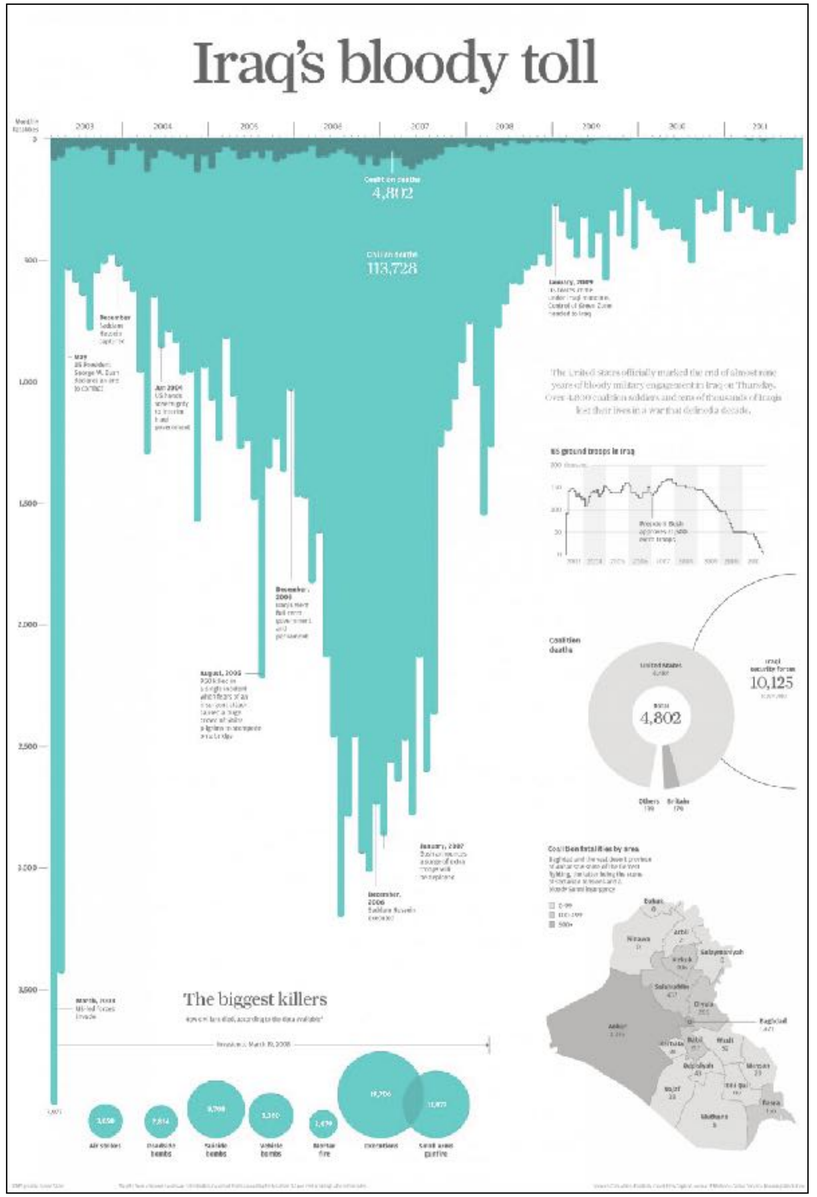
Protanope

Tritanope

*[Seriously Colorful:  Advanced Color Principles & Practices. Stone.Tableau Customer Conference 2014.]*

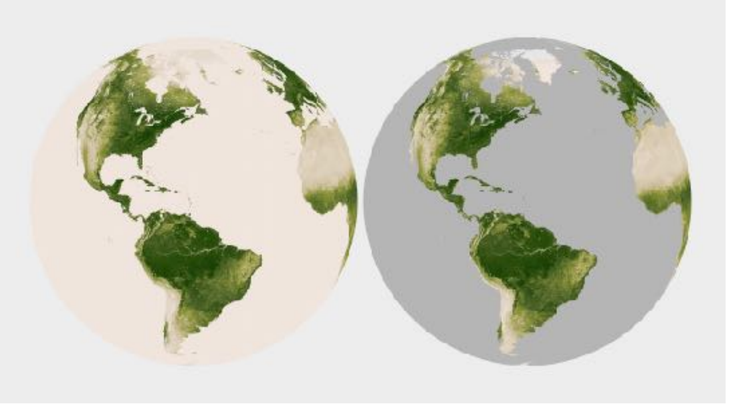
# Semantics of color

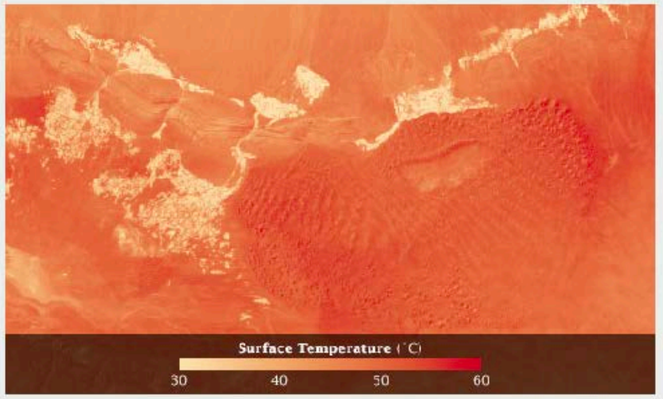


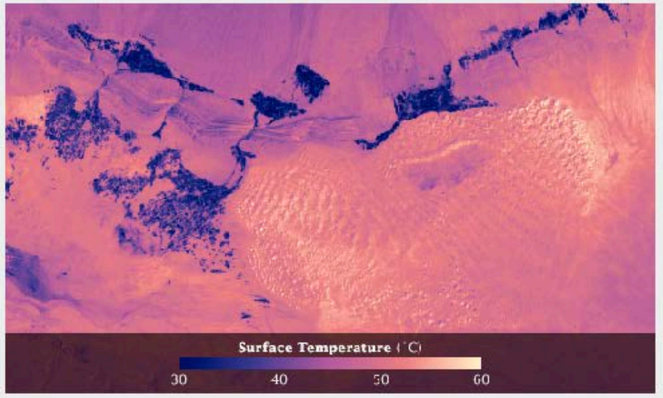



# Some advices

- Contrast is necessary to show detail
- Use perceptually linear scales
- Don’t use a rainbow color scale
- Match data type with the correct scale
- Colour deficiency and people who print papers
- Never use a rainbow color scale
- Make missing data recognisable
- Connect colour to meaning

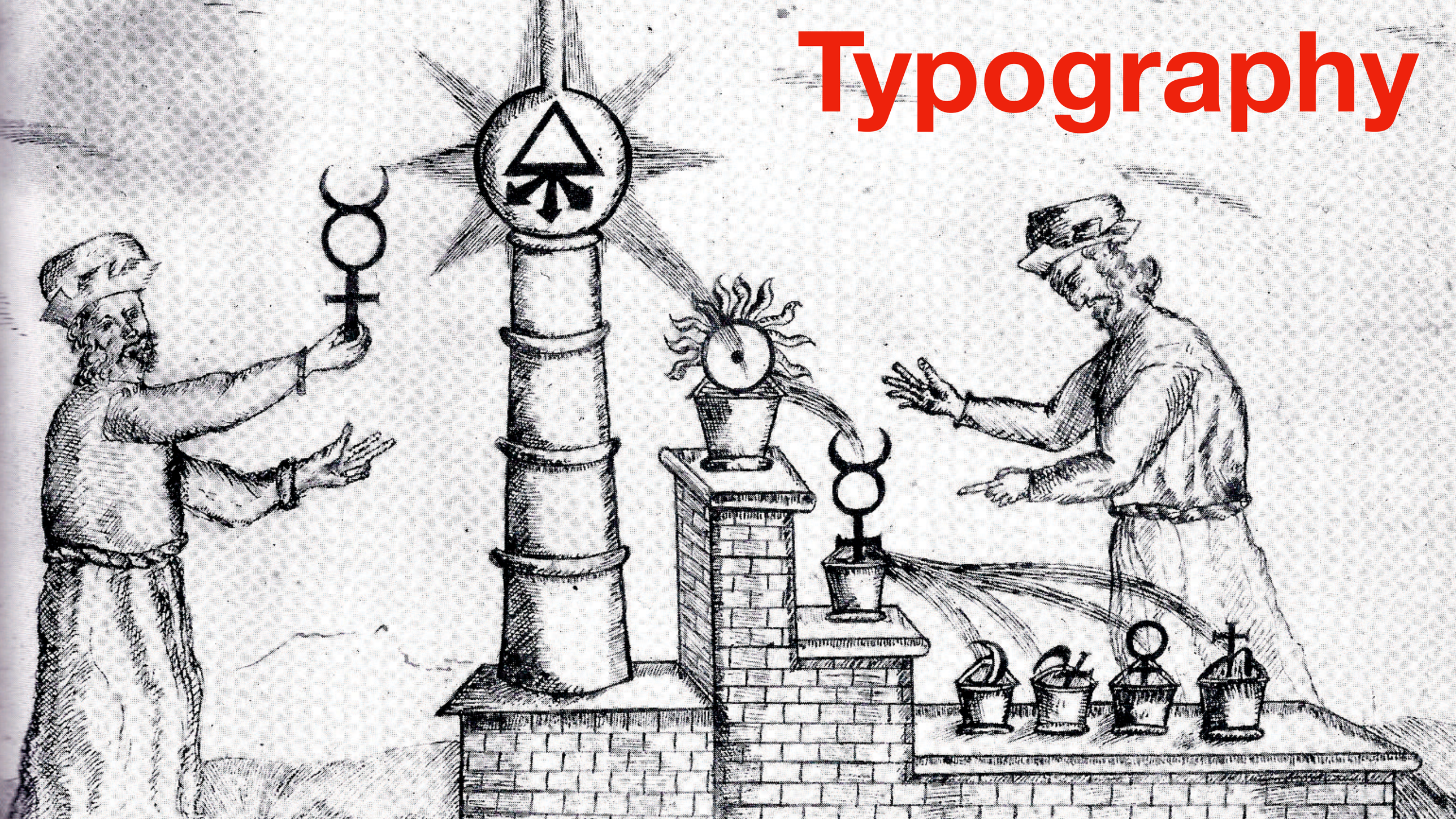





Robert Simmon, 2013

# Typography



MENÚ
DEGUSTACIÓ DE
CAÇA
Tots els DIUMENGES
13.00h-15.30h
Fred i calent de Perdiu amb
brots tendres i ceba cruixent
Arròs de Llebre i bolets
Guatlles en escabetx
Filet de Cérvol amb salsa
de porto i madurixes
Civet de Senglar amb
melmelada de nabius
Vi negre Rioja criança
Postre i café

KIDSEXCHANGE

# Typefaces

## Serif

**Bodoni**

is a beautiful, classic serif typeface

**Georgia**

is a serif designed for screen design

## Sans serif

**Helvetica**

The most popular

Open Sans

is a sans-serif specially designed for screen reading

# Typefaces are voices

*Shit!*

Shhhhh...

NATURALMENTE

Seriously

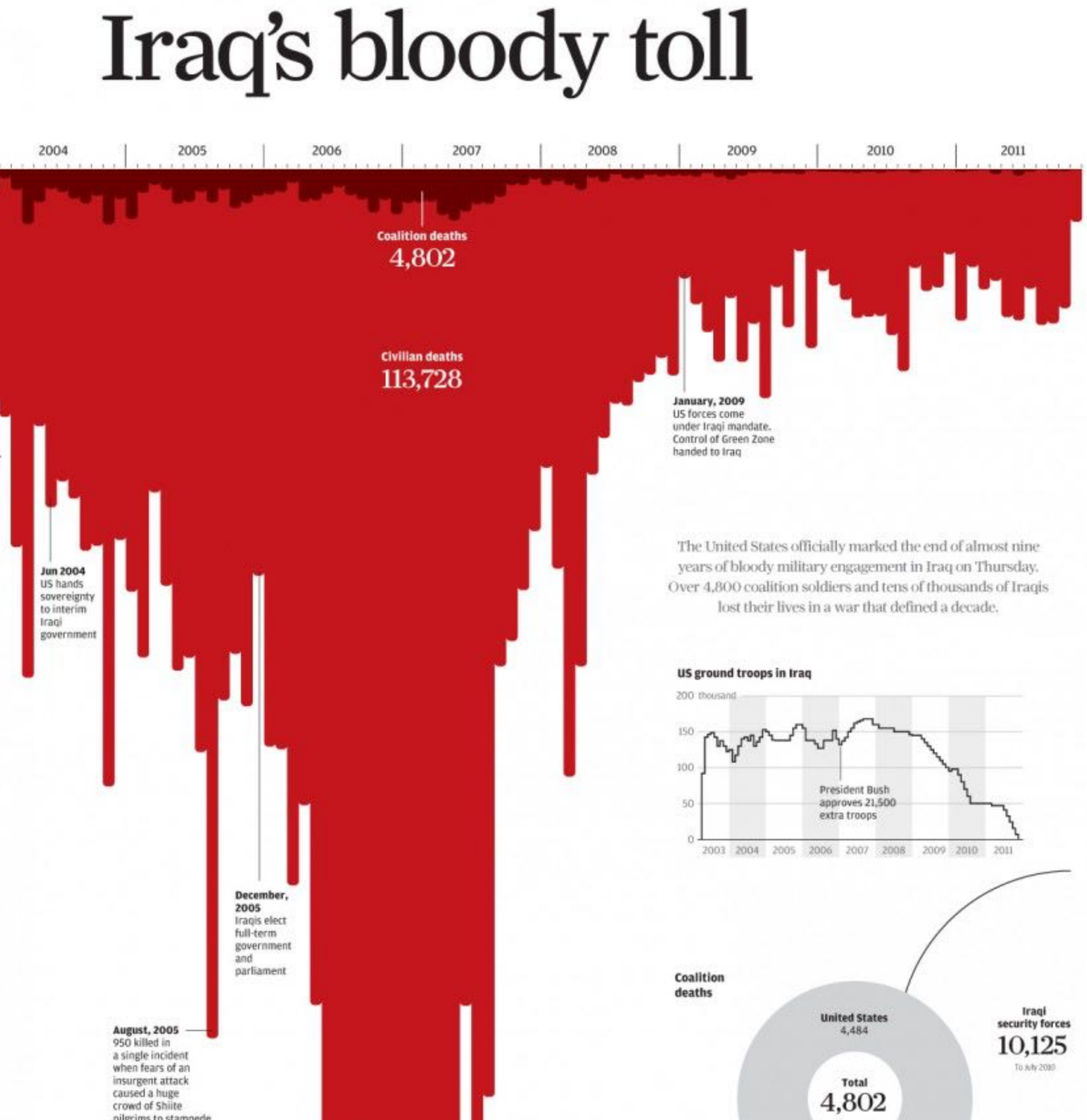
Smooth

## Make sure they don’t send the wrong message

Data visualisation

Data visualisation

Data visualisation




No more than two fonts

**Better if they** are different

Try to mix a serif **with a sans**

# Align left

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc sit amet laoreet quam, ut pellentesque nulla. Etiam pellentesque, magna ut laoreet facilisis, tellus metus consequat purus, quis vulputate lectus dolor sit amet massa. Suspendisse sollicitudin interdum pharetra. Aliquam erat volutpat. Sed sodales ullamcorper lobortis. Fusce tellus nibh, feugiat ac posuere a, sodales eleifend ligula. Ut ac dignissim felis, fermentum commodo velit. Maecenas tincidunt dapibus egestas. Curabitur malesuada semper nulla, non facilisis velit tristique ultrices.

Ut ultricies nunc in felis ultrices, vitae pharetra mi feugiat. Praesent ut turpis ac nibh dictum vulputate. Phasellus malesuada erat purus, et dapibus nunc congue dignissim. Orci varius natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Integer tempus ornare ipsum, id ornare mauris iaculis quis. Nulla facilisi. Sed in venenatis tellus. Duis placerat finibus odio, et mollis est aliquam aliquet. Suspendisse consequat sollicitudin eros sed euismod. Suspendisse potenti.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc sit amet laoreet quam, ut pellentesque nulla. Etiam pellentesque, magna ut laoreet facilisis, tellus metus consequat purus, quis vulputate lectus dolor sit amet massa. Suspendisse sollicitudin interdum pharetra. Aliquam erat volutpat. Sed sodales ullamcorper lobortis. Fusce tellus nibh, feugiat ac posuere a, sodales eleifend ligula. Ut ac dignissim felis, fermentum commodo velit. Maecenas tincidunt dapibus egestas. Curabitur malesuada semper nulla, non facilisis velit tristique ultrices.

Ut ultricies nunc in felis ultrices, vitae pharetra mi feugiat. Praesent ut turpis ac nibh dictum vulputate. Phasellus malesuada erat purus, et dapibus nunc congue dignissim. Orci varius natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Integer tempus ornare ipsum, id ornare mauris iaculis quis. Nulla facilisi. Sed in venenatis tellus. Duis placerat finibus odio, et mollis est aliquam aliquet. Suspendisse consequat sollicitudin eros sed euismod. Suspendisse potenti.

# Mind the gaps

It is better not to justify the text, it modifies the kerning between words and characters, and affects readability.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc sit amet laoreet quam, ut pellentesque nulla. Etiam pellentesque, magna ut laoreet facilisis, tellus metus consequat purus, quis vulputate lectus dolor sit amet massa. Suspendisse sollicitudin interdum pharetra. Aliquam erat volutpat. Sed sodales ullamcorper lobortis. Fusce tellus nibh, feugiat ac posuere a, sodales eleifend ligula. Ut ac dignissim felis, fermentum commodo velit. Maecenas tincidunt dapibus egestas. Curabitur malesuada semper nulla, non facilisis velit tristique ultrices.

MEGAFLICKS

# Save other alignments

For titles, for example. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

# Line length

## ~28 spaces

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc sit amet laoreet quam, ut pellentesque nulla. Etiam pellentesque, magna ut laoreet facilisis, tellus metus consequat purus, quis vulputate lectus dolor sit amet massa. Suspendisse sollicitudin interdum pharetra. Aliquam erat volutpat. Sed sodales ullamcorper lobortis. Fusce tellus nibh, feugiat ac posuere a, sodales eleifend ligula. Ut ac dignissim felis, fermentum commodo velit. Maecenas tincidunt dapibus egestas. Curabitur malesuada semper nulla, non facilisis velit tristique ultrices.

## ~75 spaces (including blank)

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc sit amet laoreet quam, ut pellentesque nulla. Etiam pellentesque, magna ut laoreet facilisis, tellus metus consequat purus, quis vulputate lectus dolor sit amet massa. Suspendisse sollicitudin interdum pharetra. Aliquam erat volutpat. Sed sodales ullamcorper lobortis. Fusce tellus nibh, feugiat ac posuere a, sodales eleifend ligula. Ut ac dignissim felis, fermentum commodo velit. Maecenas tincidunt dapibus egestas. Curabitur malesuada semper nulla, non facilisis velit tristique ultrices.

## Too long

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc sit amet laoreet quam, ut pellentesque nulla. Etiam pellentesque, magna ut laoreet facilisis, tellus metus consequat purus, quis vulputate lectus dolor sit amet massa. Suspendisse sollicitudin interdum pharetra. Aliquam erat volutpat. Sed sodales ullamcorper lobortis. Fusce tellus nibh, feugiat ac posuere a, sodales eleifend ligula. Ut ac dignissim felis, fermentum commodo velit. Maecenas tincidunt dapibus egestas. Curabitur malesuada semper nulla, non facilisis velit tristique ultrices.

# Line spacing

The distance from the baseline of one line of type to another is called *line spacing*. It is also called *leading*, in reference to the strips of lead used to separate lines of metal type. The default setting in most layout and imaging software is 120 percent of the type size. Thus 10-pt type is set with 12 pts of line spacing. Designers play with line spacing in order to create distinctive layouts. Reducing the standard distance creates a denser typographic color—while risking collisions between ascenders and descenders.

6/6 SCALA PRO
(6 pt type with 6 pts line spacing, or "set solid")

The distance from the baseline of one line of type to another is called *line spacing*. It is also called *leading*, in reference to the strips of lead used to separate lines of metal type. The default setting in most layout and imaging software is 120 percent of the type size. Thus 10-pt type is set with 12 pts of line spacing. Designers play with line spacing in order to create distinctive layouts. Reducing the standard distance creates a denser typographic color—while risking collisions between ascenders and descenders.

6/7.2 SCALA PRO
(Auto spacing; 6 pt type with 7.2 pts line spacing)

The distance from the baseline of one line of type to another is called *line spacing*. It is also called *leading*, in reference to the strips of lead used to separate lines of metal type. The default setting in most layout and imaging software is 120 percent of the type size. Thus 10-pt type is set with 12 pts of line spacing. Designers play with line spacing in order to create distinctive layouts. Reducing the standard distance creates a denser typographic color—while risking collisions between ascenders and descenders.

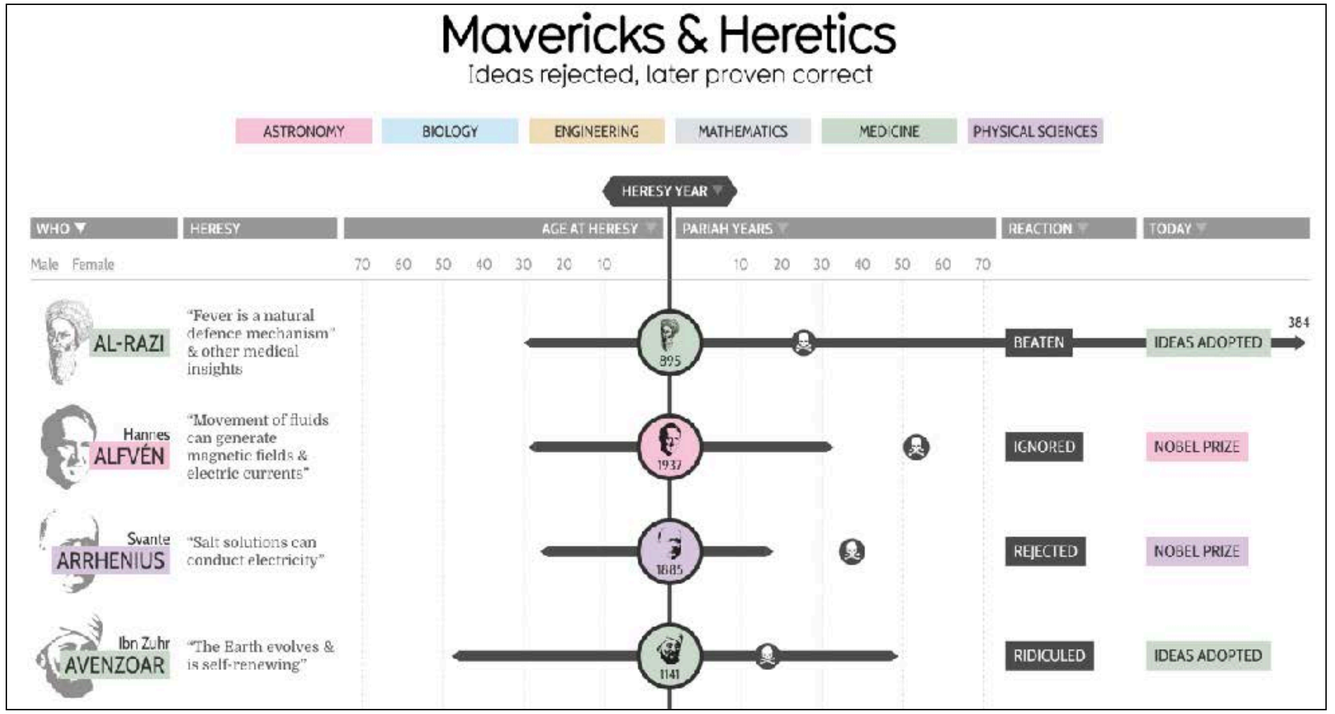
6/8 SCALA PRO
(6 pt type with 8 pts line spacing)

The distance from the baseline of one line of type to another is called *line spacing*. It is also called *leading*, in reference to the strips of lead used to separate lines of metal type. The default setting in most layout and imaging software is 120 percent of the type size. Thus 10-pt type is set with 12 pts of line spacing. Designers play with line spacing in order to create distinctive layouts. Reducing the standard distance creates a denser typographic color—while risking collisions between ascenders and descenders.

6/12 SCALA PRO
(6 pt type with 12 pts line spacing)

# Hierarchy

information is beautiful
About Blog Data Books Work With Us Training Jobs Contact
Senseless
lines / labels
Deaths per Hour
Years of Conflict (Log)
0 3 4 5 10 30
0
10
Insurgency in Aceh
Gaza-Israel conflict
War in Af
Lebanese Civil War
Francisco Macías Nguema
Second Boer War
Congo War
Bosnian War
Iraq War
Idi Amin
First Sudanese Civil War
Charles Taylor (Liberia)
Eritrean War of Independence
Gulf War
Maji-Maji Rebellion
Syrian Civil War
Second Sudanese Civil War
Bosnian Genocide
Russo-Japanese War
Mexican Revolution
Iran-Iraq War


Mavericks & Heretics
Ideas rejected, later proven correct
ASTRONOMY
BIOLOGY
ENGINEERING
MATHEMATICS
MEDICINE
PHYSICAL SCIENCES
HERESY YEAR
WHO
HERESY
AGE AT HERESY
PARIAH YEARS
REACTION
TODAY
Male Female
70 60 50 40 30 20 10
10 20 30 40 50 60 70
AL-RAZI
"Fever is a natural defence mechanism" & other medical insights
895
384
BEATEN
IDEAS ADOPTED
Hannes ALFVÉN
"Movement of fluids can generate magnetic fields & electric currents"
1937
IGNORED
NOBEL PRIZE
Svante ARRHENIUS
"Salt solutions can conduct electricity"
1885
REJECTED
NOBEL PRIZE
Ibn Zuhr AVENZOAR
"The Earth evolves & is self-renewing"
1141
RIDICULED
IDEAS ADOPTED

# Style

## *Italic*

Save italics for:

- Quote *text*
- Words in other *languages*
- Not *normative* words

## **Bold**

Use bold to **emphasize**

Bold **helps** creating hierarchy

**Use, don’t abuse**

## ALL CAPS

Save all caps for specific cases:

YOU DON’T WANT TO SHOUT TO THE READER!

# Font size

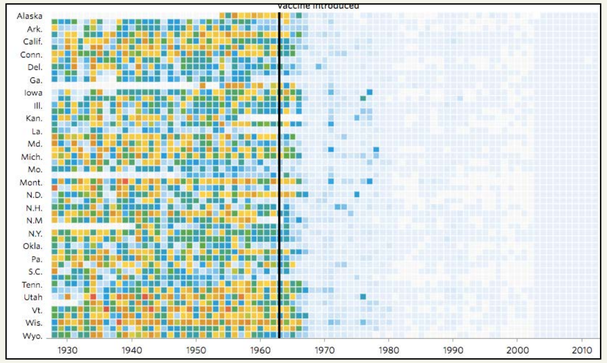
- 6-8 for footnotes
- 9-12 for text
- Depends on the typeface, intended use, and intention

# Storytelling with data

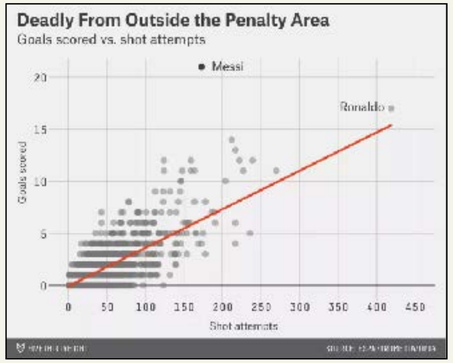
# How to tell a data story

## Step by step process

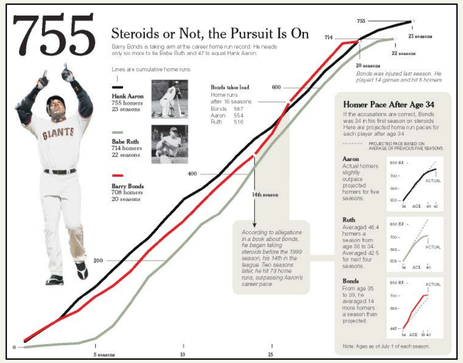
Never about the data!




Story Arc

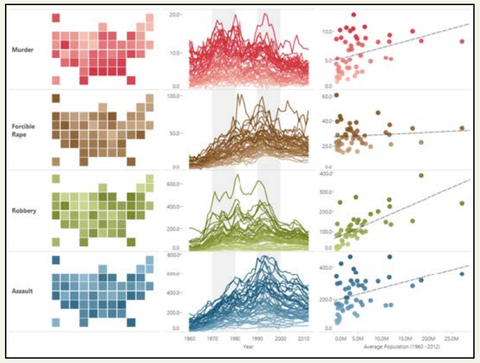



Filter



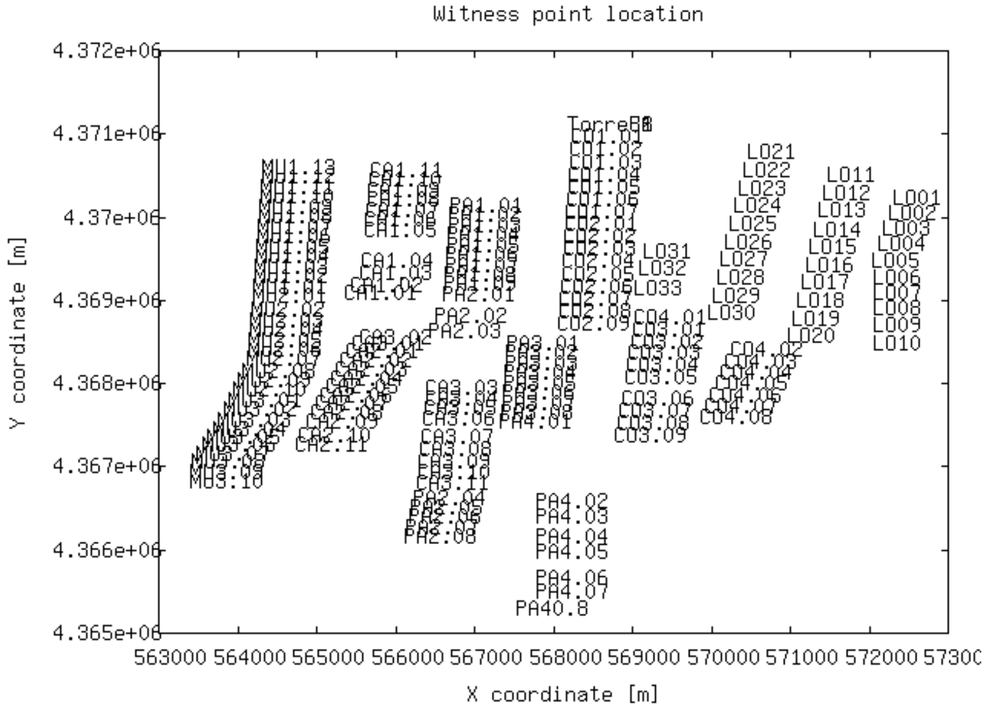

Make the right chart

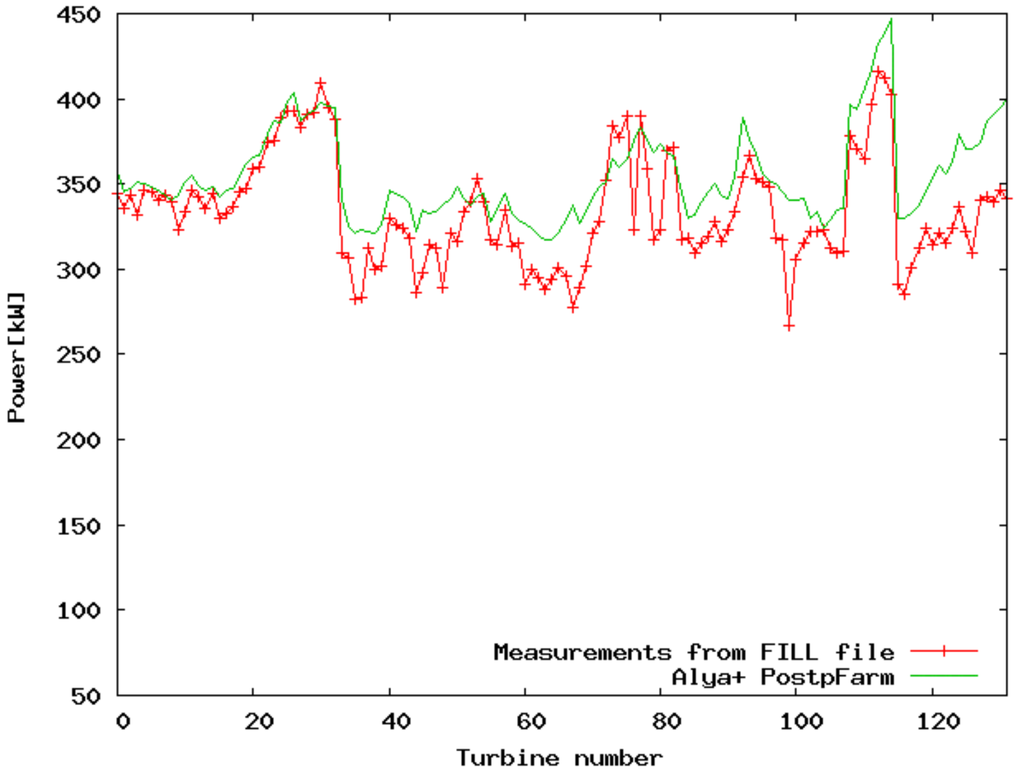
Design

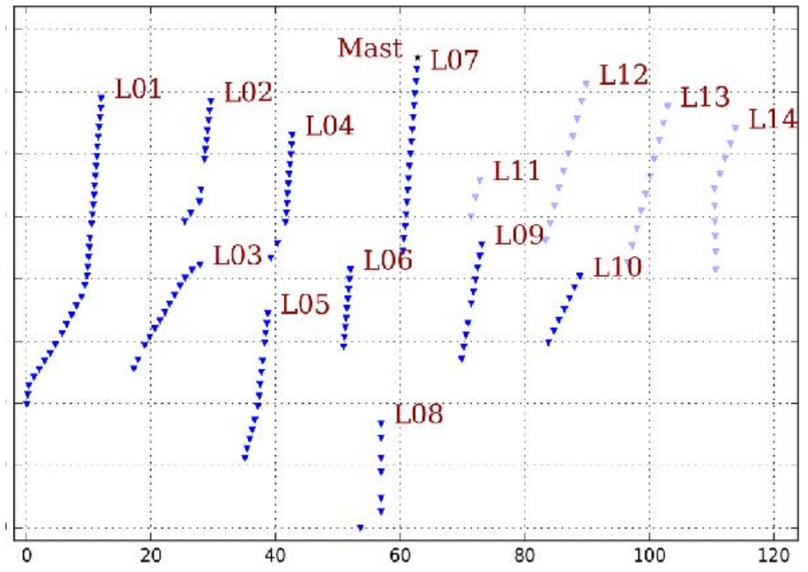


# Original plots

- Simulation of the wind to calculate energy production in a real wind farm
- Comparison of simulation results (green) vs. on-site measurements (red)
- Annual average of each individual windmill (N=132)
- Dataset has Power (kW), ID, array ID, and location

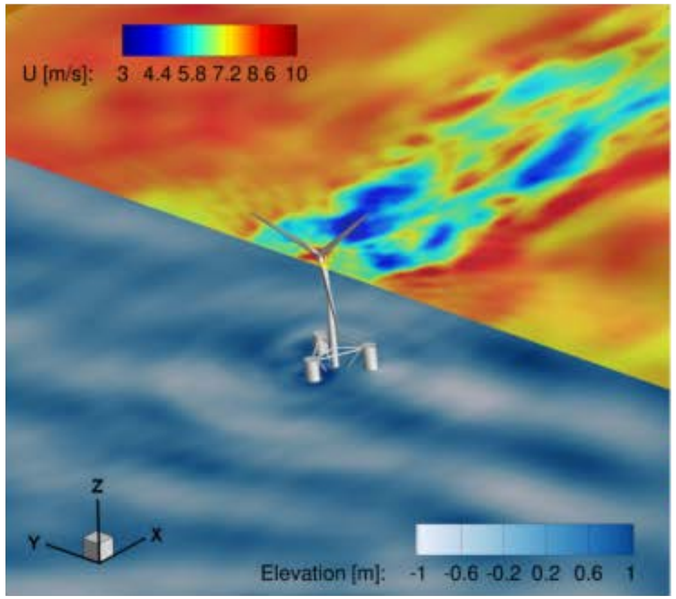


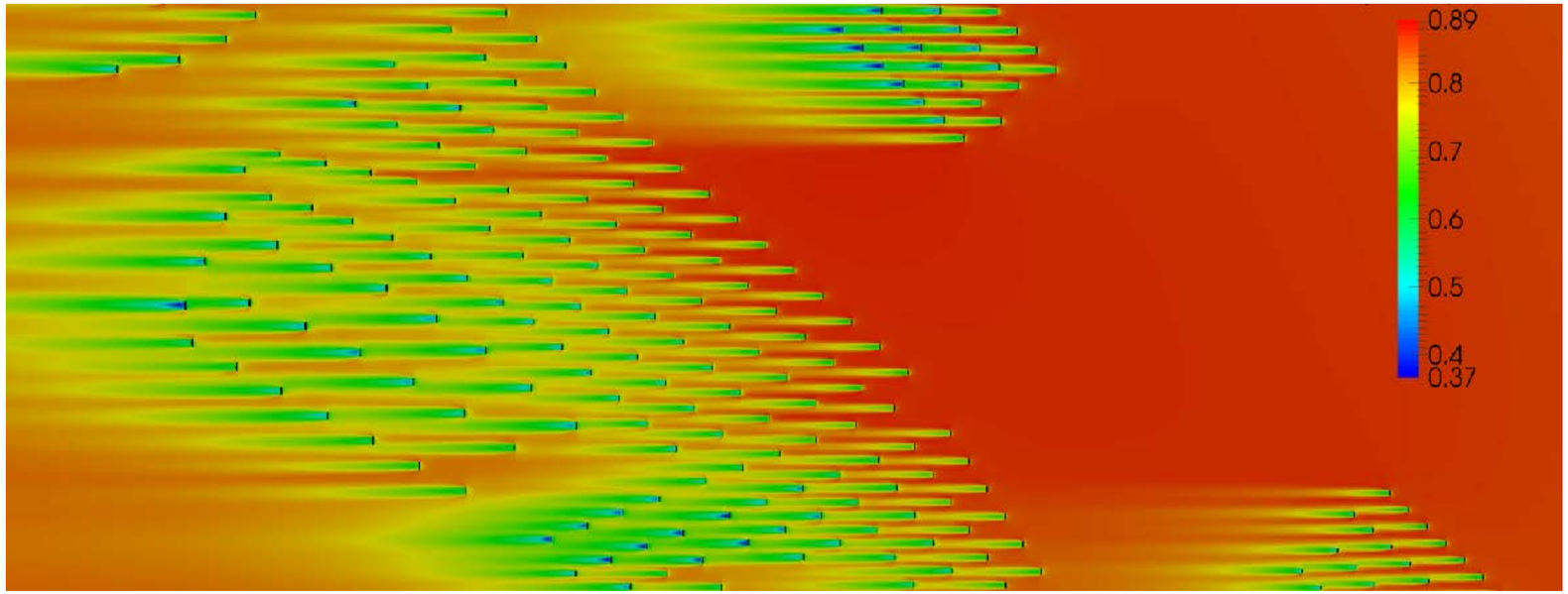



# Additional Data

- 3D Simulation of the wind velocity over complex terrain geometry
- Usually shown as 2D cuts at the centre of the blades

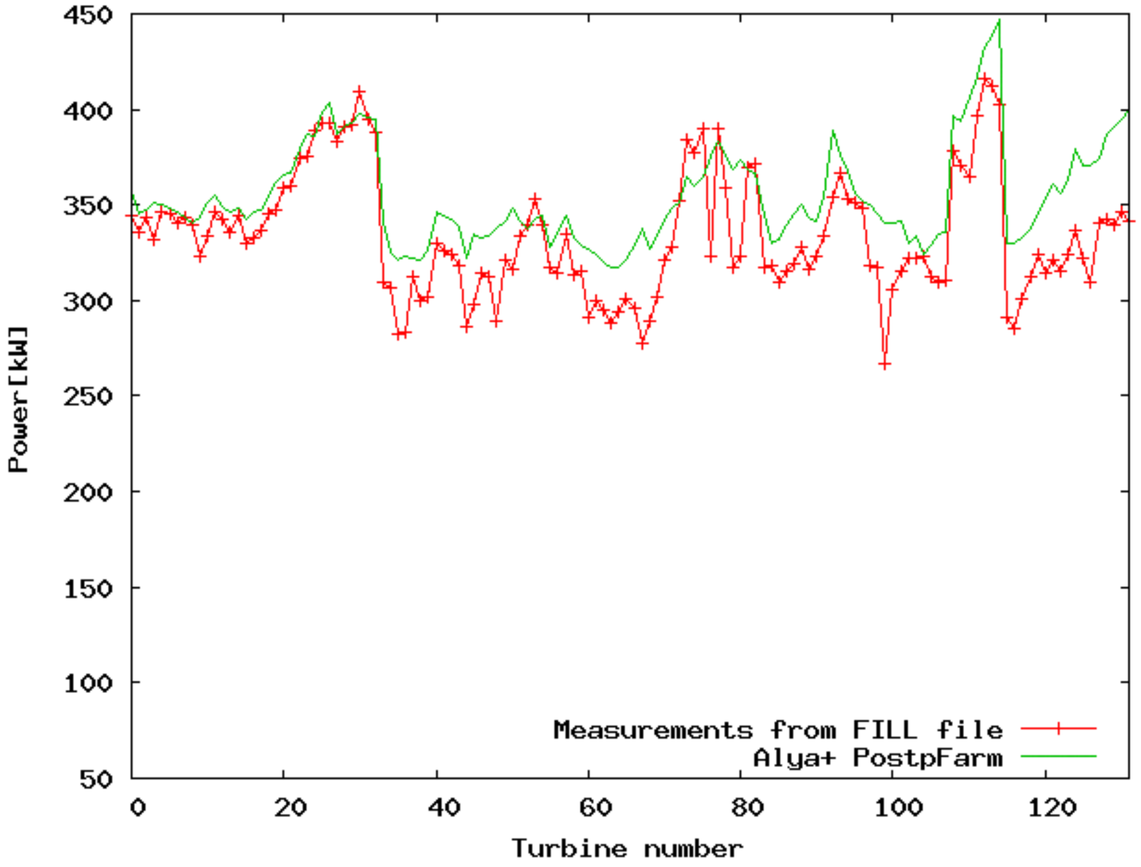



https://www.energy.gov/eere




Barcelona Supercomputing Center

# Original plot




**A basic checklist**

**Who is the audience?** Project partners in energy company

**How will it be used?** Presentation of results/dissemination poster

**What is the goal?** Compare results vs. measurements
Assess the accuracy of the model

**What is the challenge?** Stress the differences between values and clarify what is represented on each axis

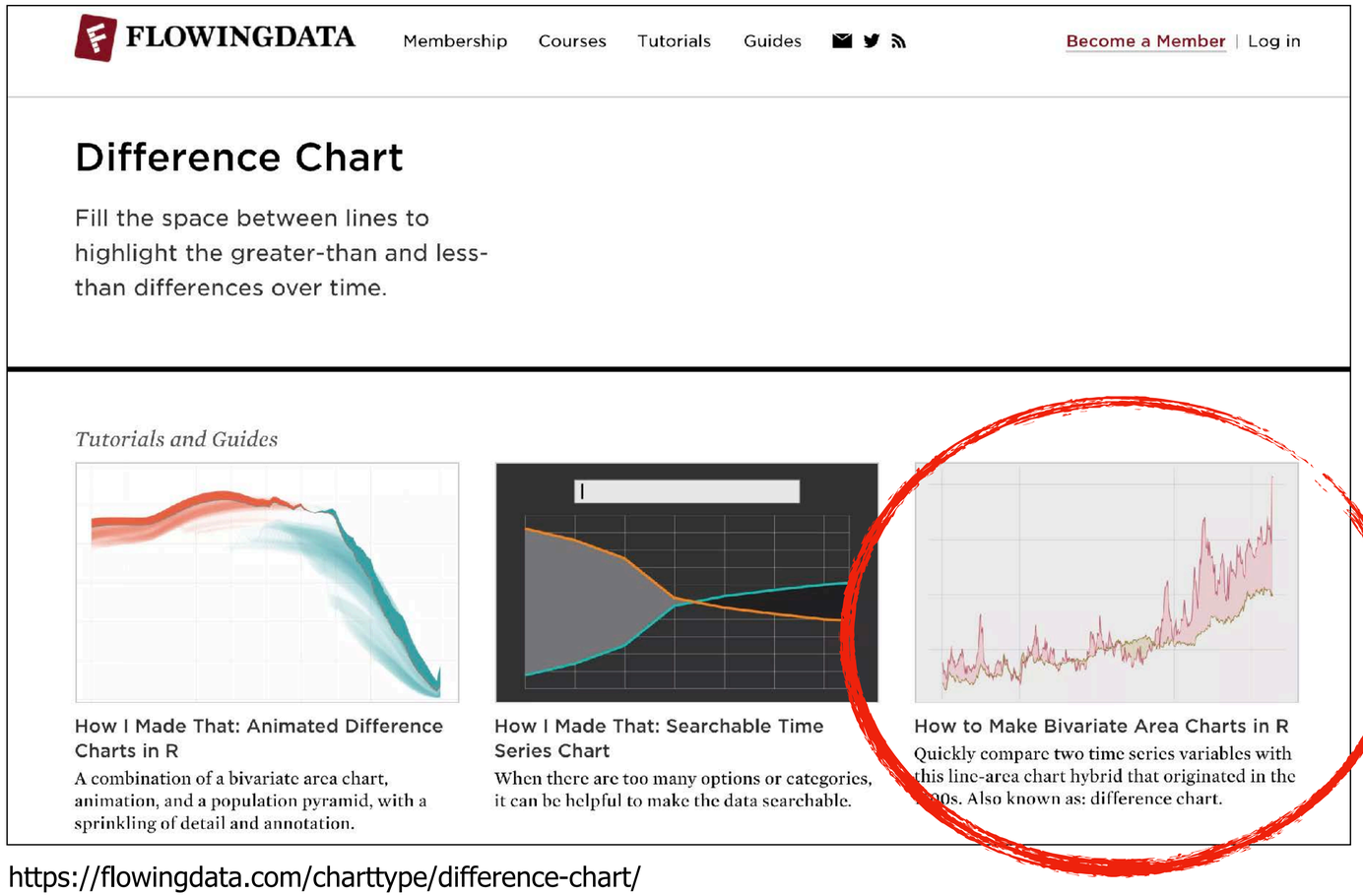
# Difference Chart

## A.K.A. Bivariate Area Charts

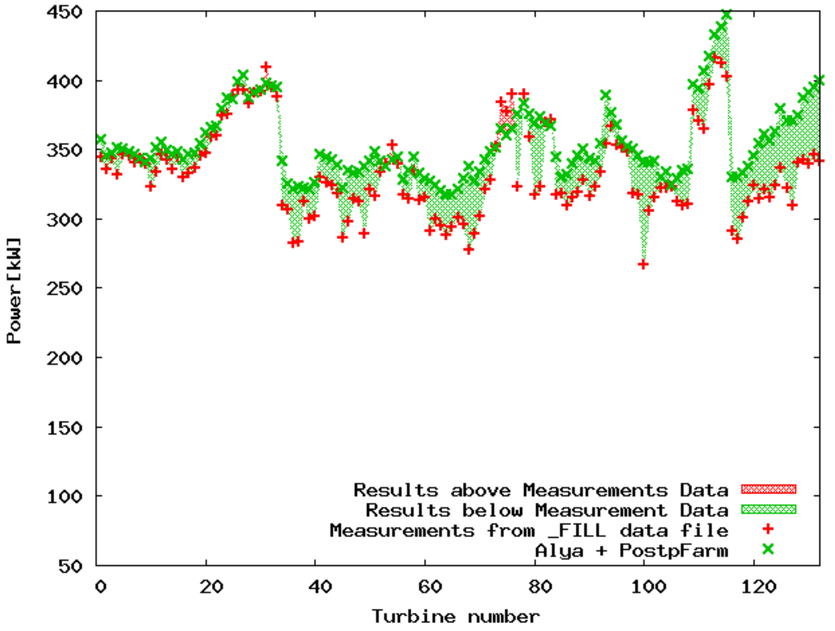
Displays two lines and the area between them.

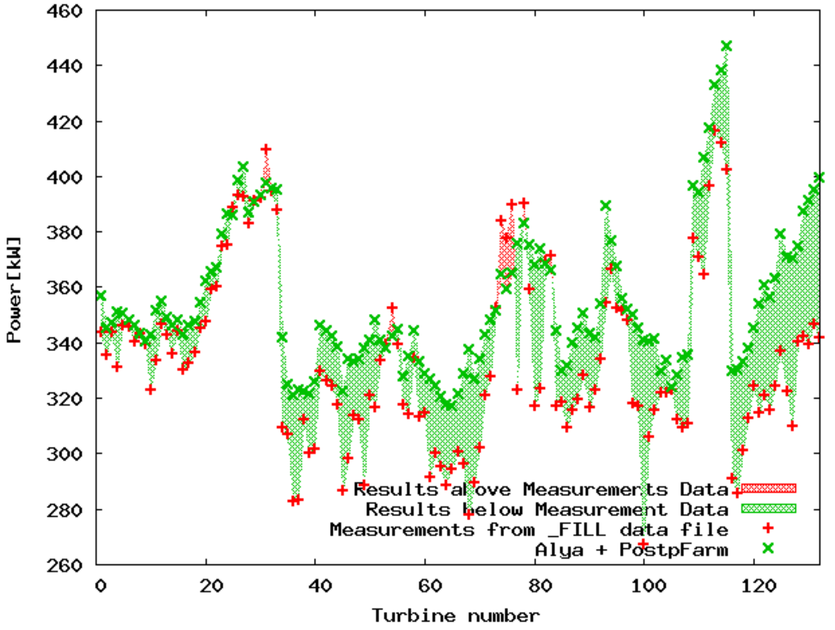
Different colour for positive and negative difference

Mostly used for continuous data, usually time series



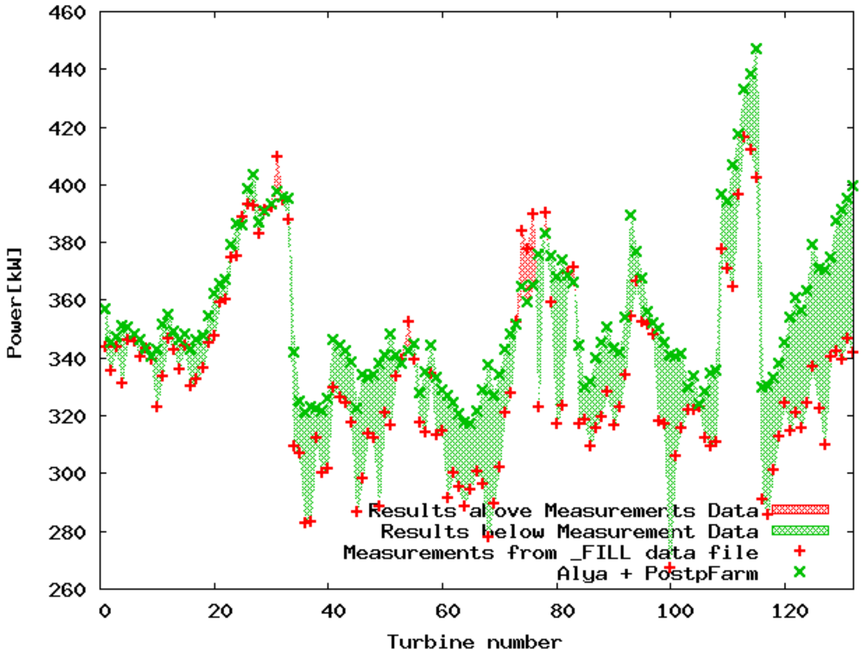

https://flowingdata.com/charttype/difference-chart/

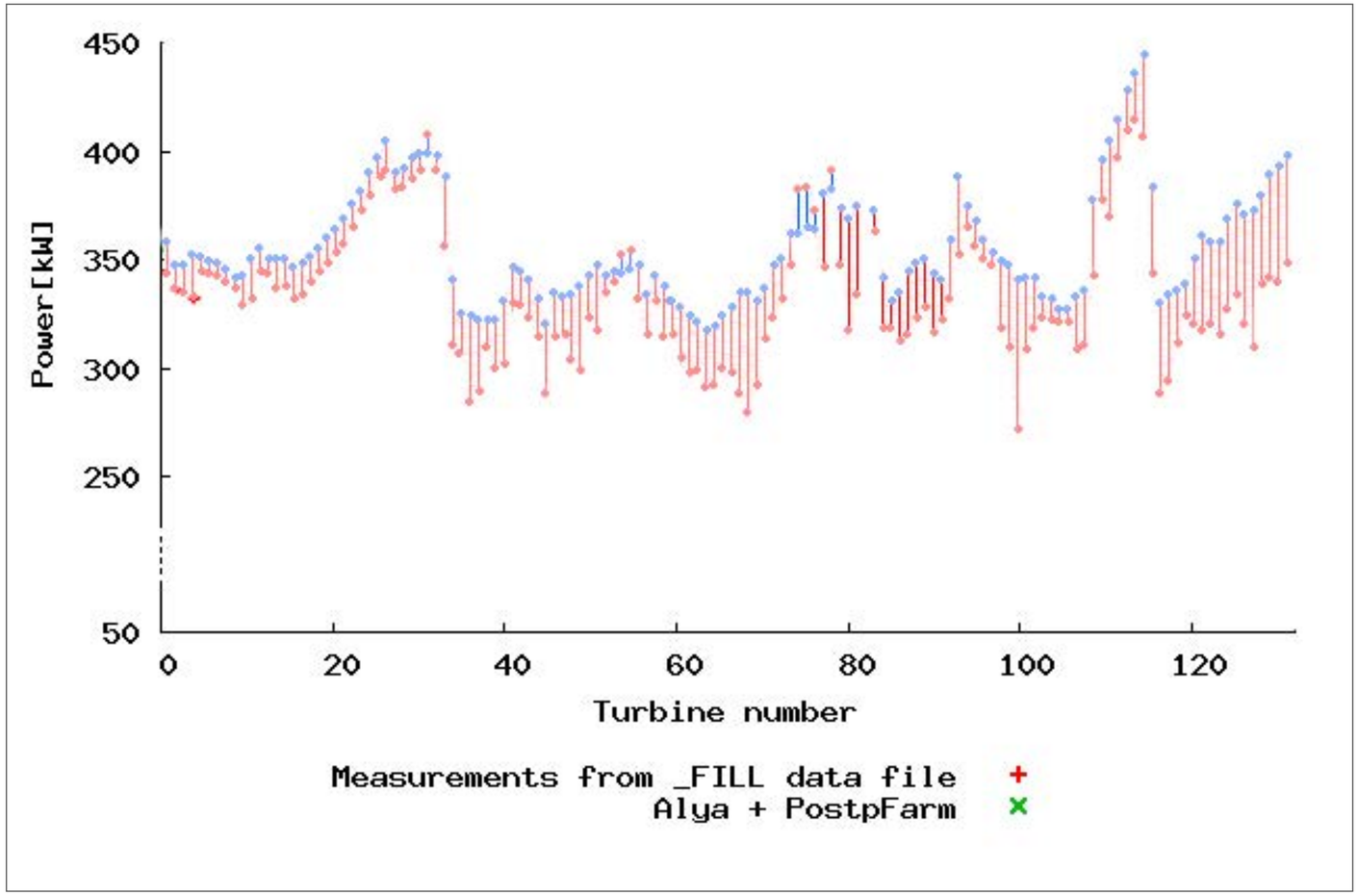


## First prototype: Fake difference chart with separated data points

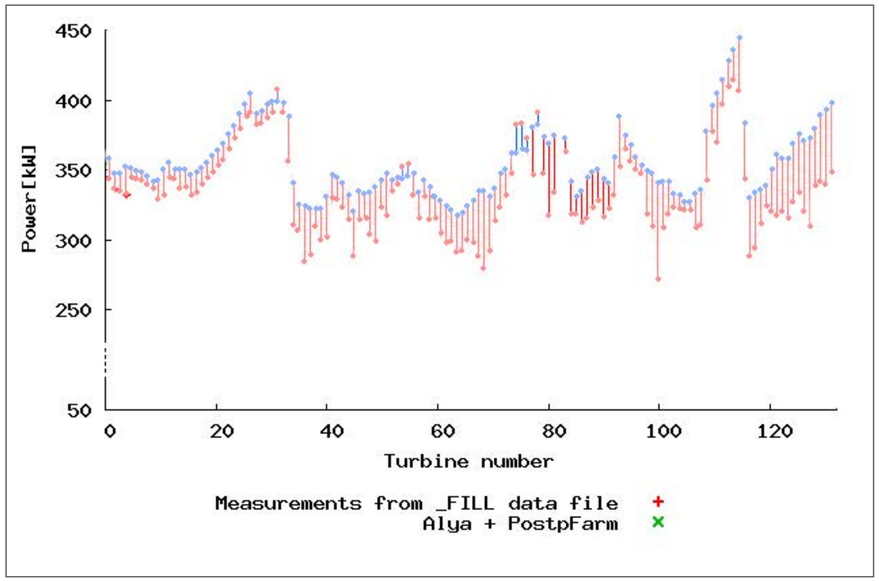
- Hand made in Adobe Illustrator
- Test different colour schemes, avoid “traffic light” colors
- Break the perceived continuity of the standard line chart

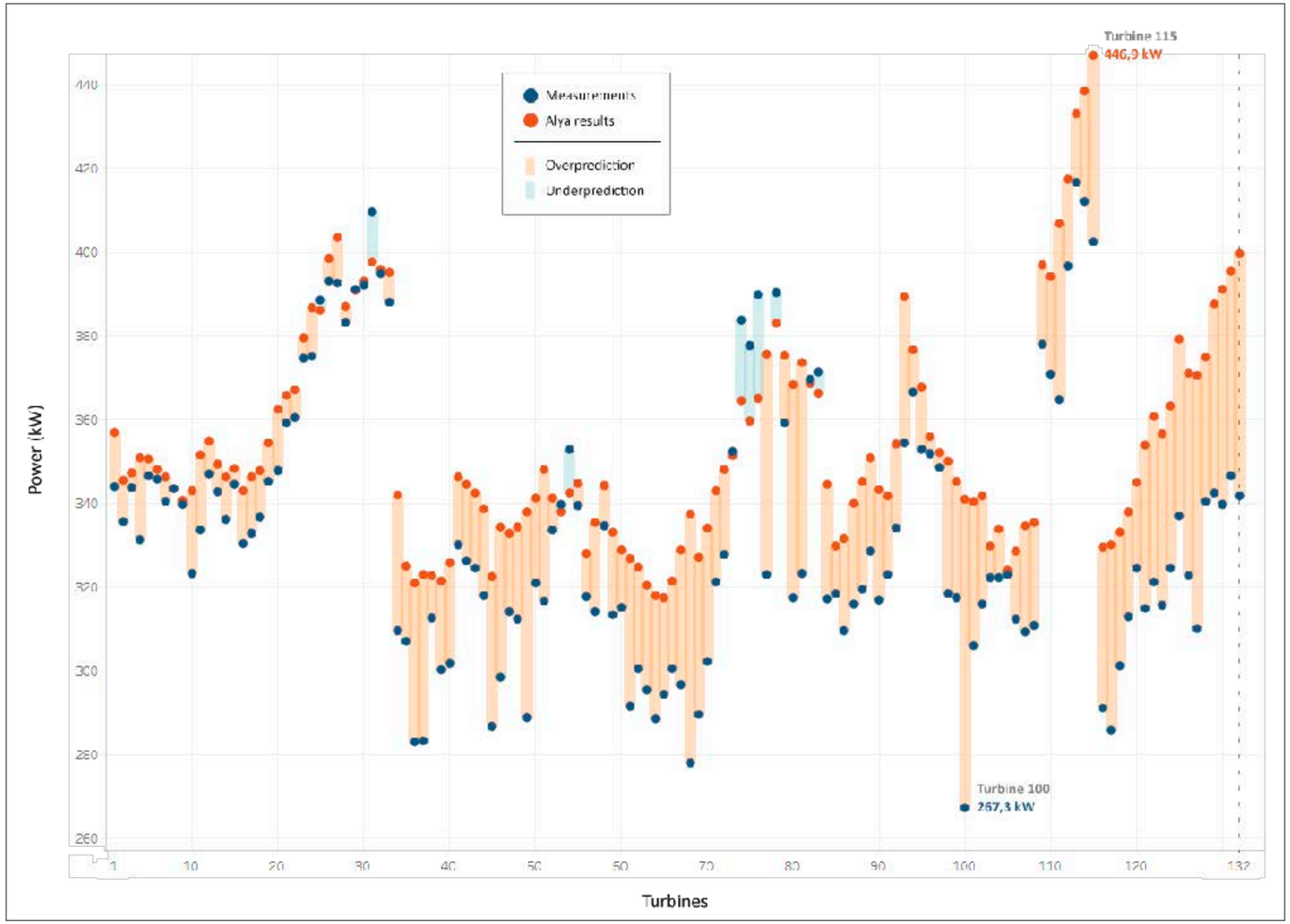


# Second prototype: Real data in Tableau Public and export to Illustrator

- Choose your tool
- Re-adjust colours (dots darker than lines)
- **Export SVG** to design software
- Add annotations, create legend box, tweak axes. Any task not conditioned by the number of elements



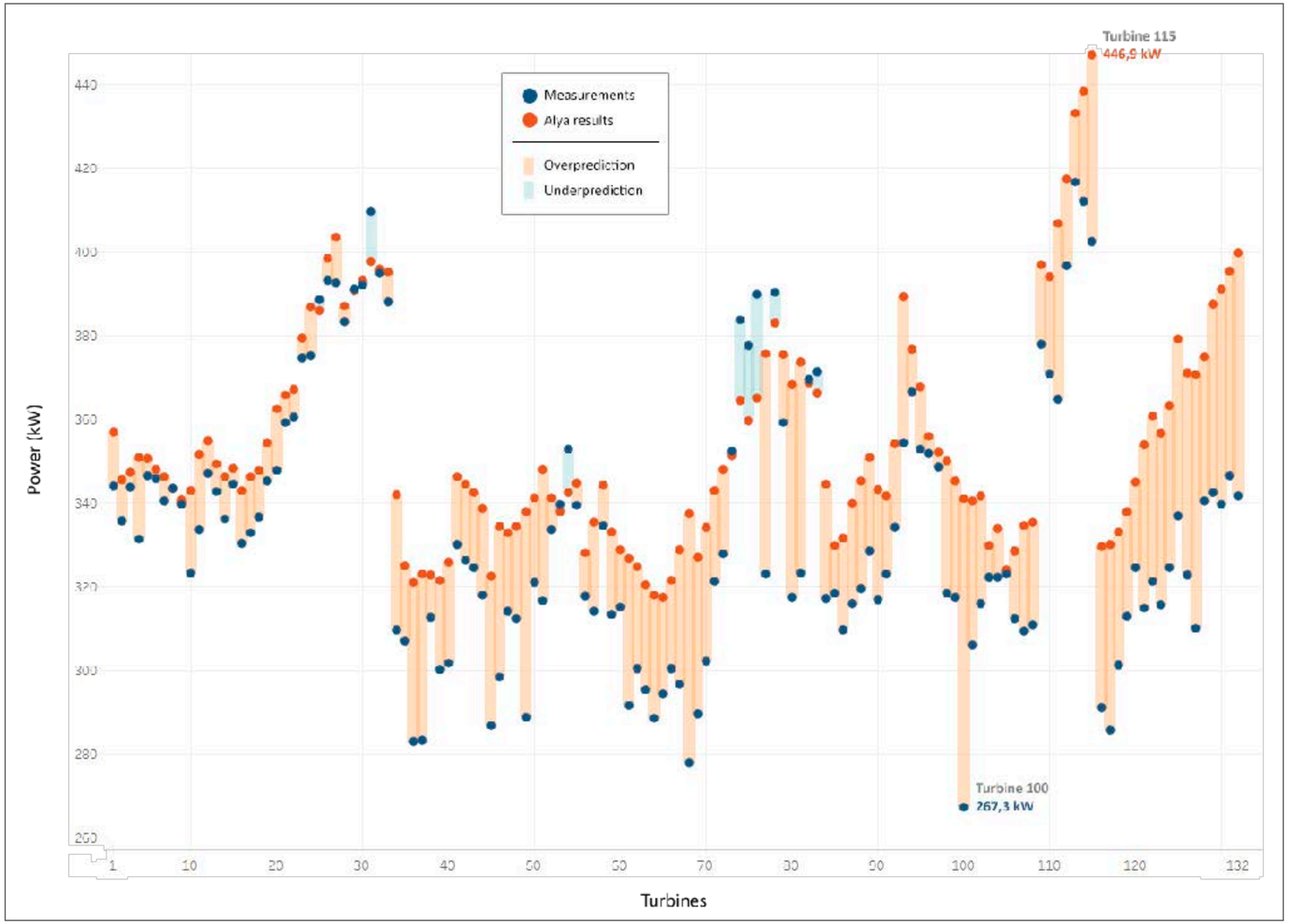



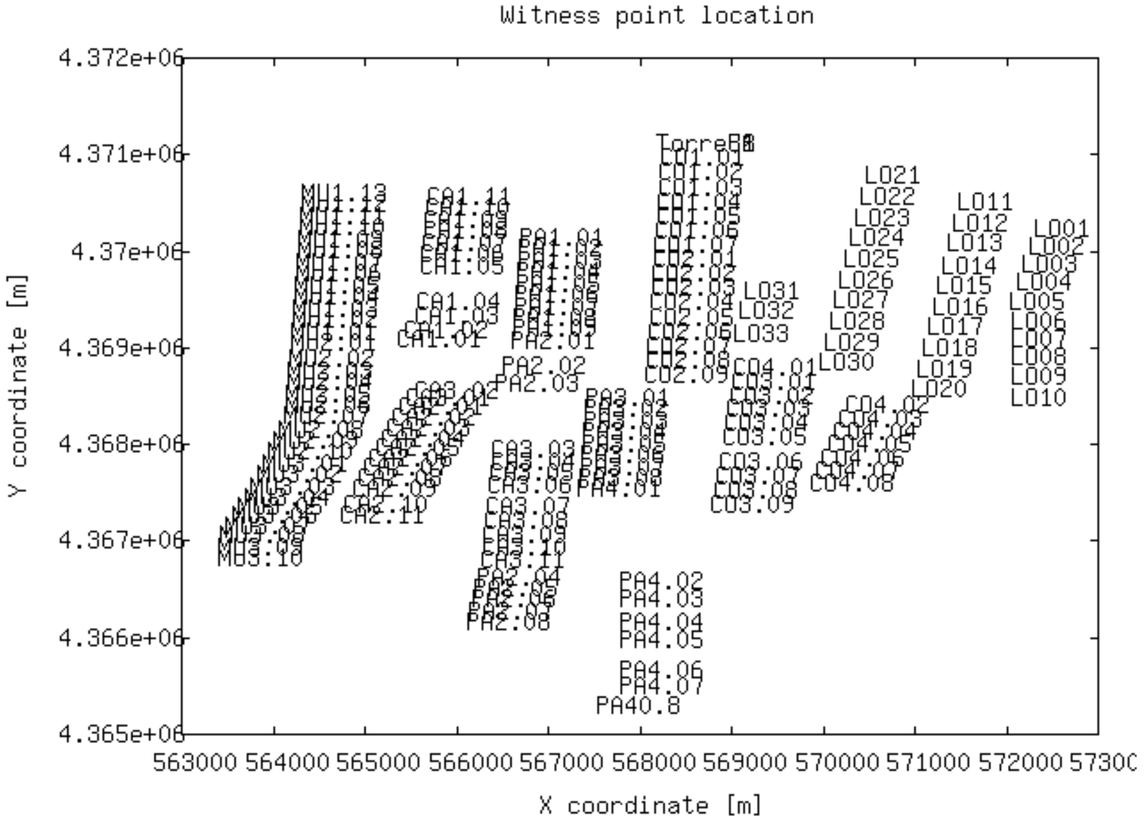
# Second prototype: Real data in Tableau Public

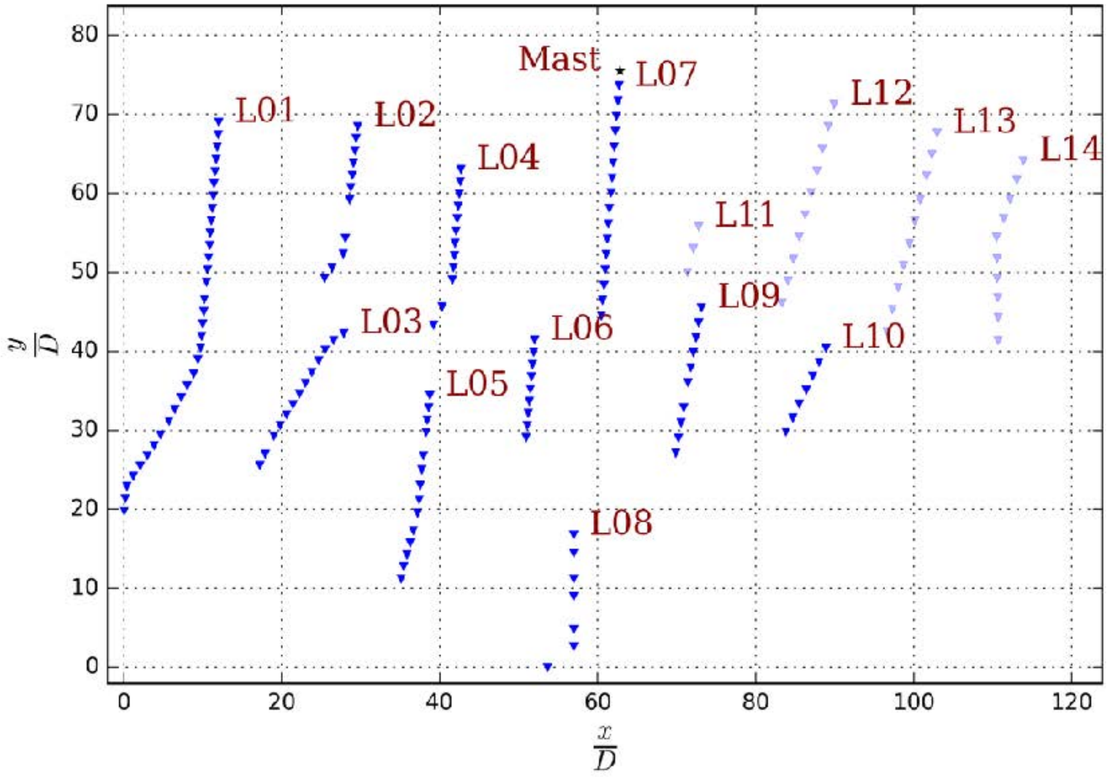
**What is the challenge?**

Stress the differences between values and clarify what is represented on each axis

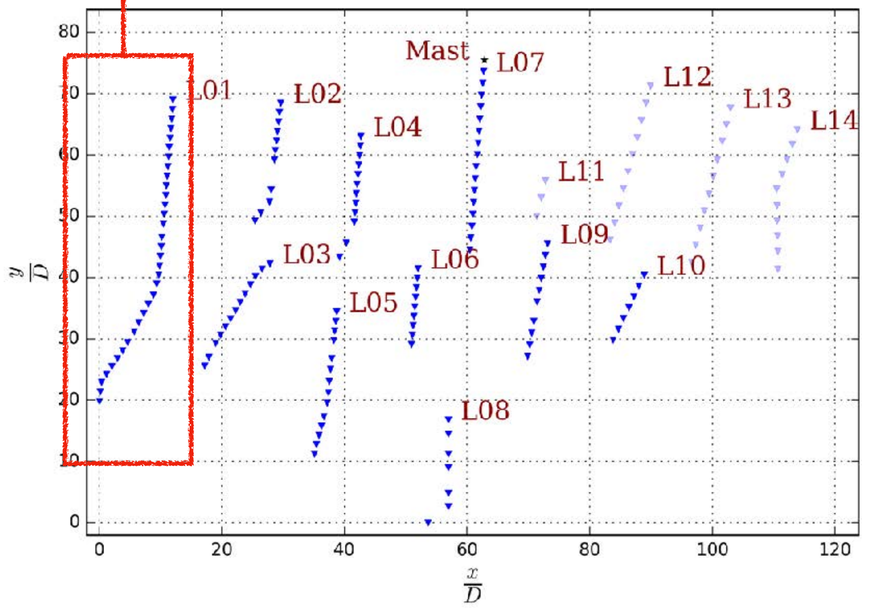
A continuous axis does not work for turbine number: associated to continuous data

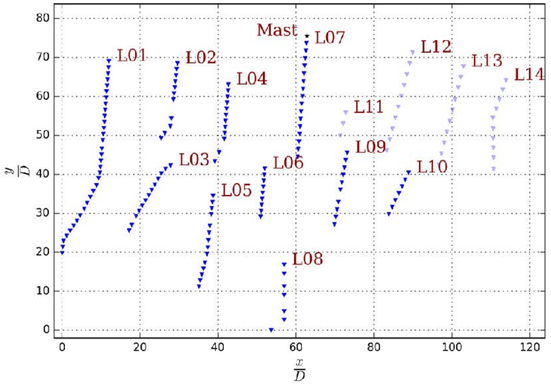



Witness point location
Y coordinate [m]
X coordinate [m]
4.372e+06
4.371e+06
4.37e+06
4.369e+06
4.368e+06
4.367e+06
4.366e+06
4.365e+06
563000 564000 565000 566000 567000 568000 569000 570000 571000 572000 5730C

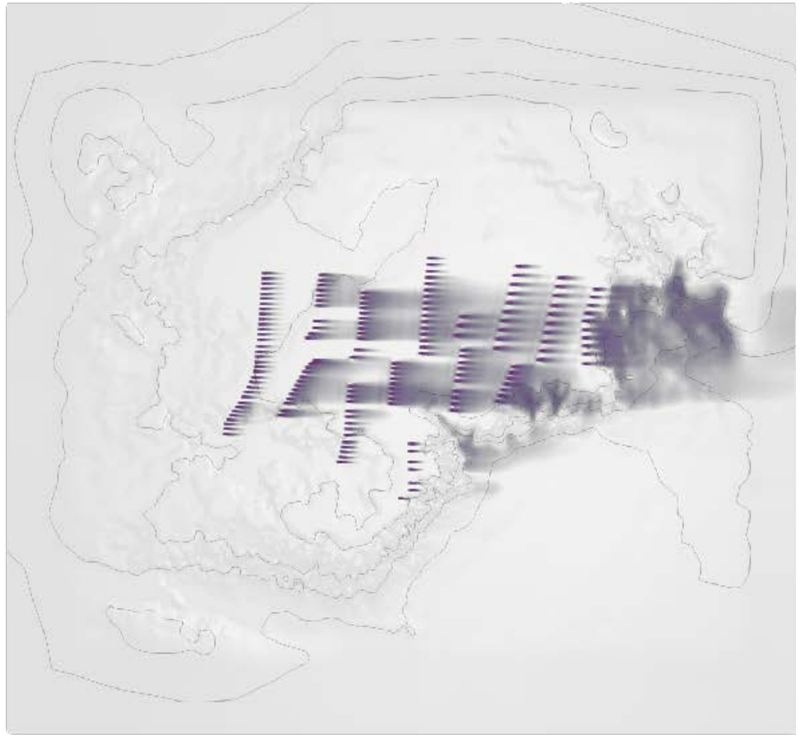

Mast
L01
L02
L03
L04
L05
L06
L07
L08
L09
L10
L11
L12
L13
L14
$\frac{y}{D}$
$\frac{x}{D}$
0 10 20 30 40 50 60 70 80
0 20 40 60 80 100 120

- Link both plots to get rid of X axis
- Break continuity while still allowing to compare differences
- Exploring the relation between physical location and power generation can give further insight

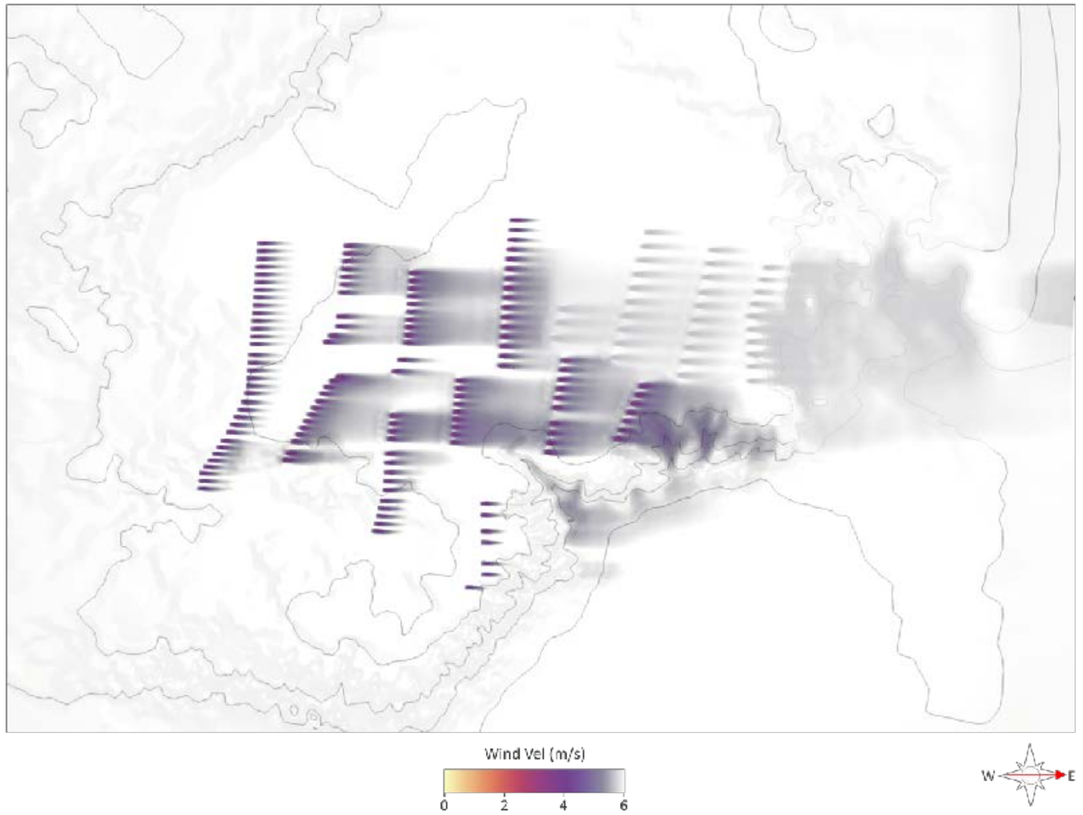




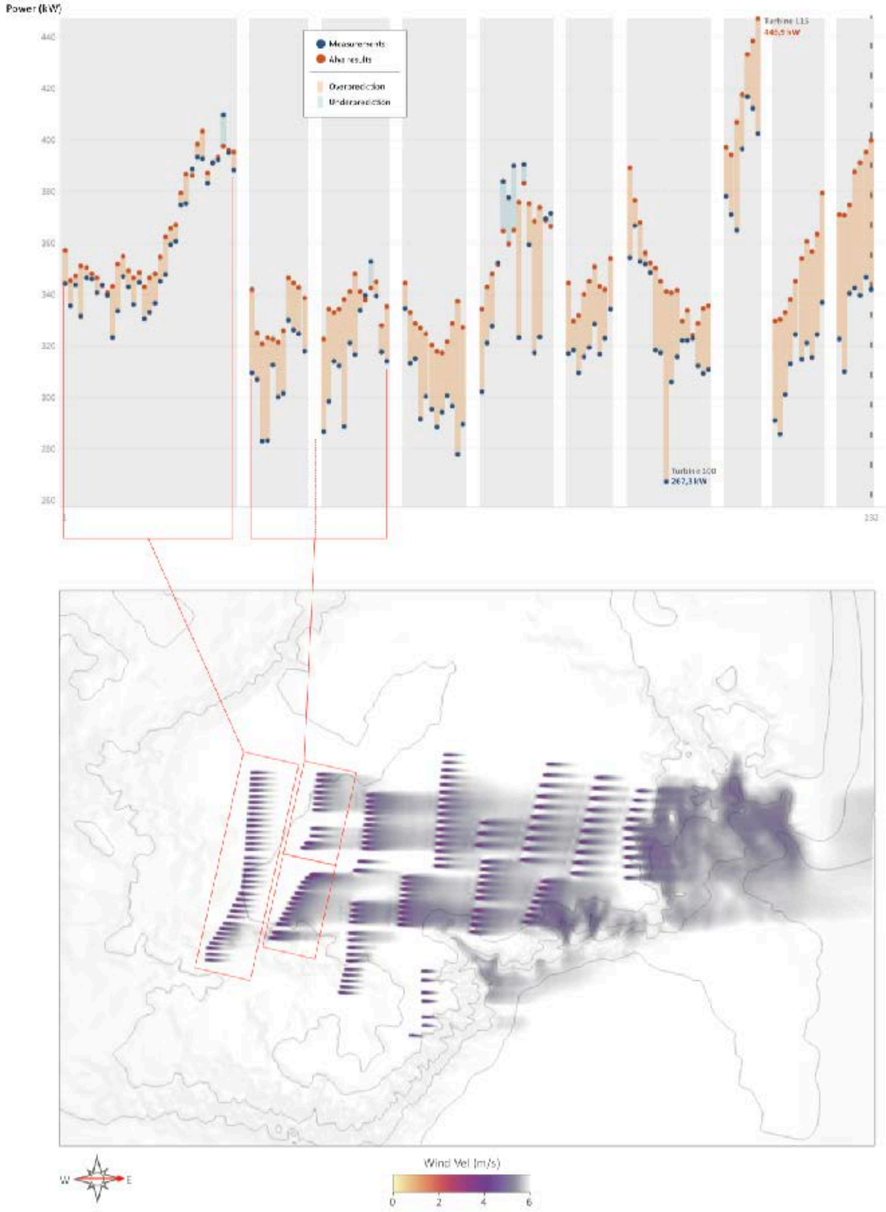

Original from Paraview (Sci-Viz software)



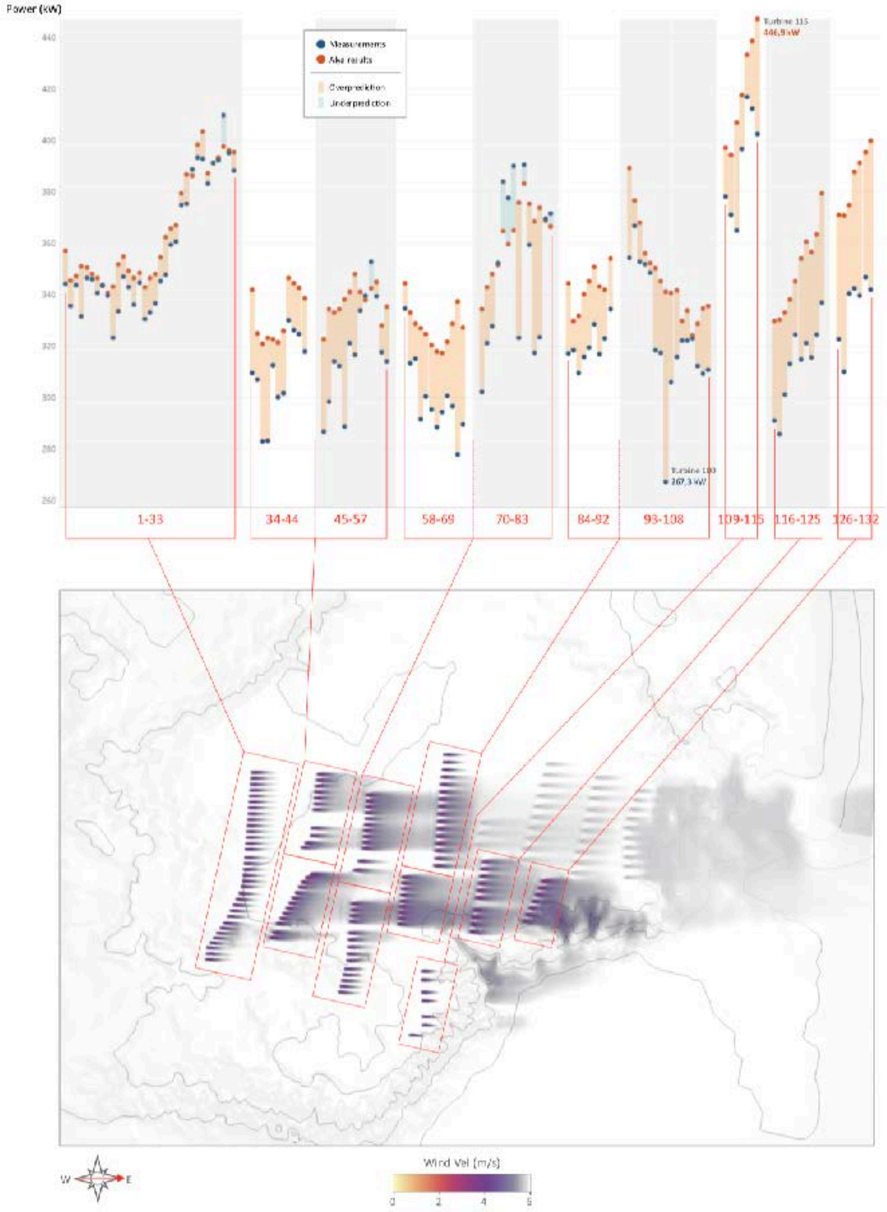
With color scale and orientation from Illustrator
Correct background brightness to make data clearer
Attenuate windmills excluded from the study



## V1

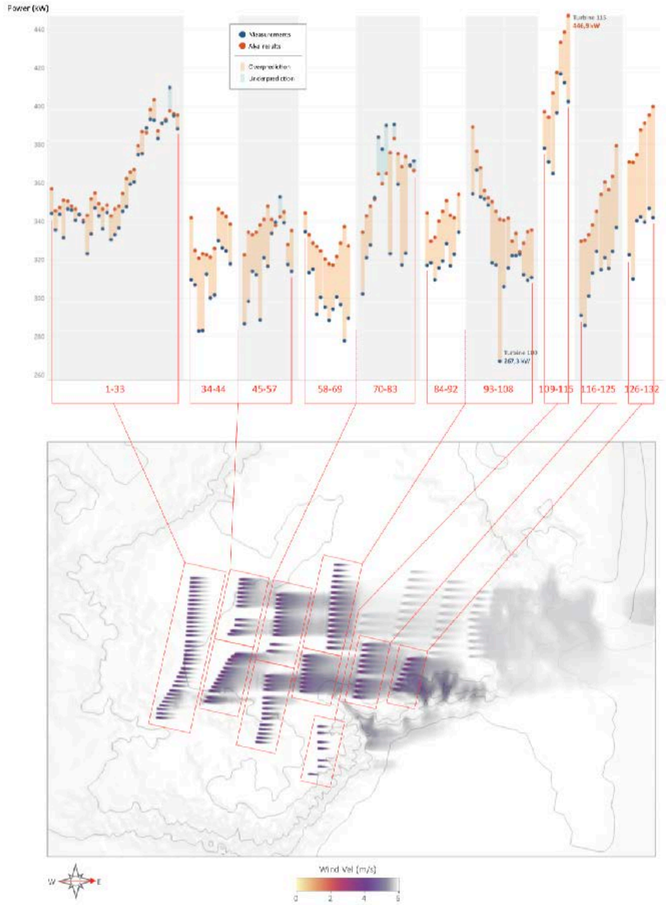

Power (kW)
Measurements
Aki results
Overprediction
Underprediction
Turbine 115
446,9 kW
Turbine 100
267,3 kW
W
E
Wind Vel (m/s)
0
2
4
6

## V2

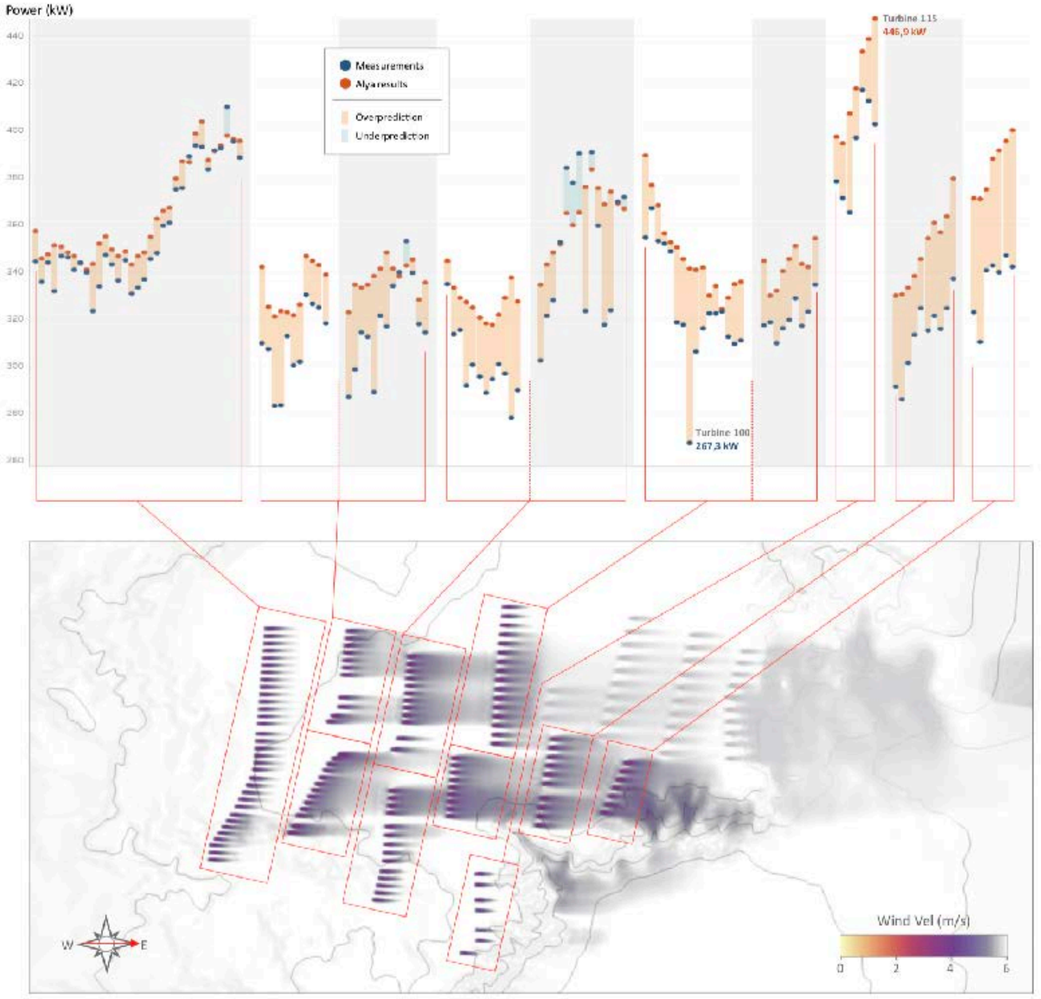

Power (kW)
Measurements
Aki results
Overprediction
Underprediction
Turbine 115
446,9 kW
Turbine 100
267,3 kW
1-33
34-44
45-57
58-69
70-83
84-92
93-108
109-115
116-125
126-132
W
E
Wind Vel (m/s)
0
2
4
6

- Cut unnecessary space of the map
- make plot narrower for composition
- remove numbers
- SWITCH group of windmills to match reading order in the map
- Legends inside the plots

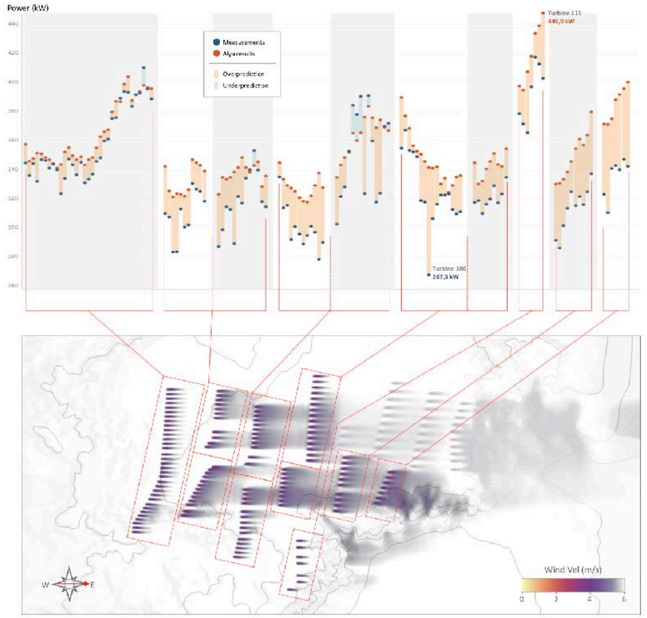
**V2**




**V4**

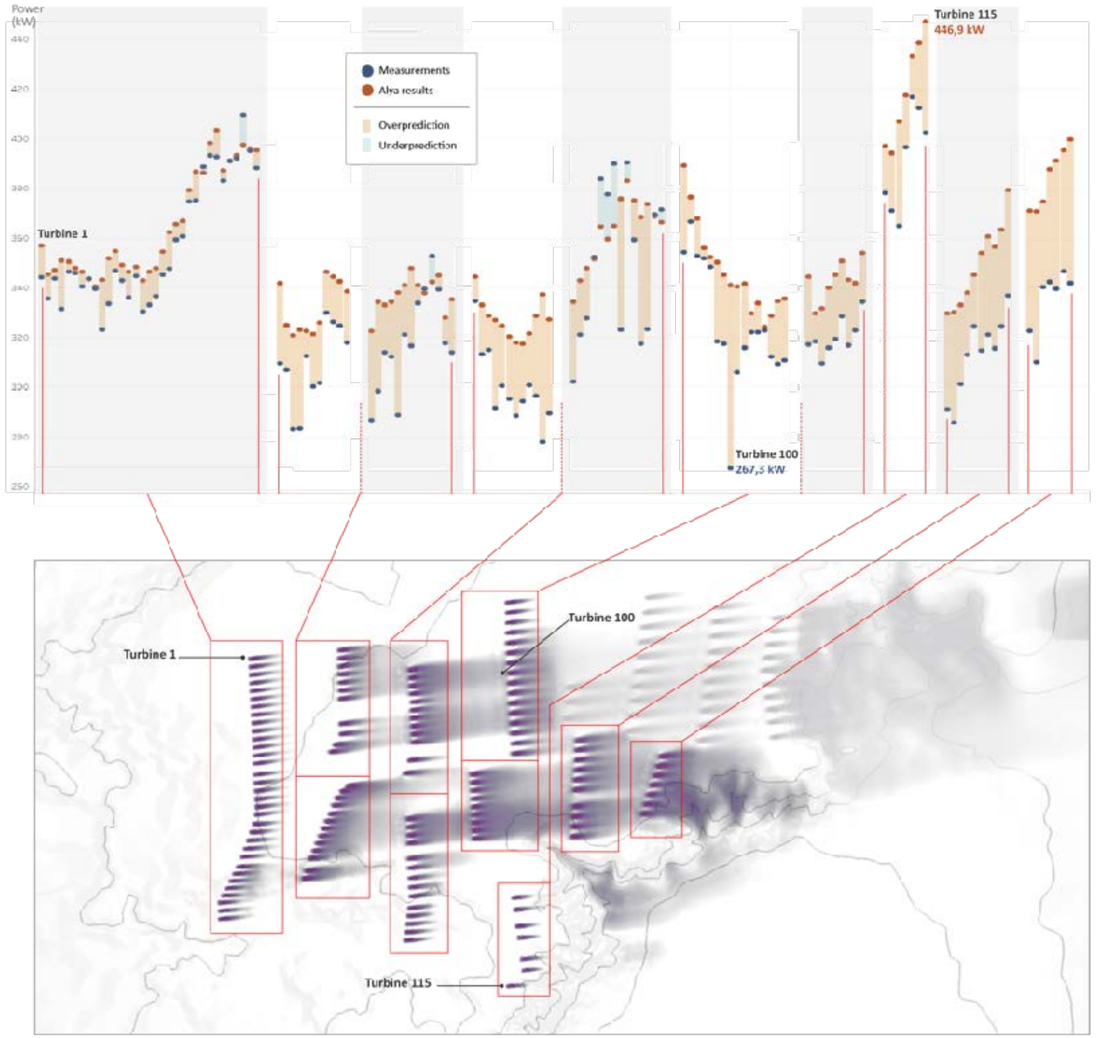


## V4




- Rotate map to put boxes straight
- Put the red brackets where the X axis was
- Mark the first, highest, and lowest-value windmills in both charts to mark reading order

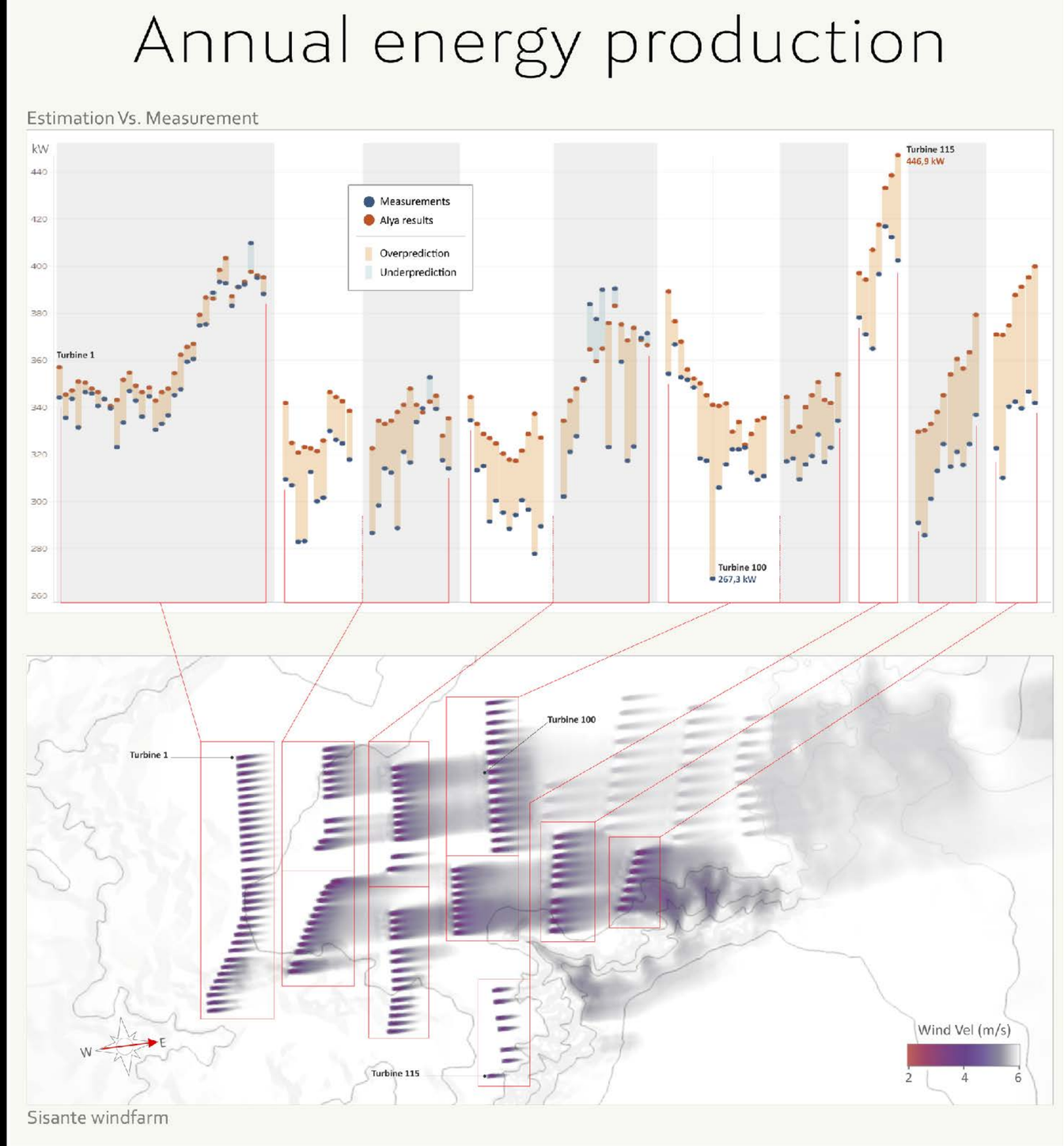
## V7



## Final image

- Enclose and separate the plots for readability: it is not a single plot (difference chart bigger, it has more information)
- Light color background to highlight the plots
- Two fonts: titles (big and light) and legends (small and bold)
- All data points against light, desaturated backgrounds for contrast

- **Unanticipated perk of this plot**: Relation between turbulence and the accuracy of the results. Known effect by researchers but a good way to explain it to non-experts

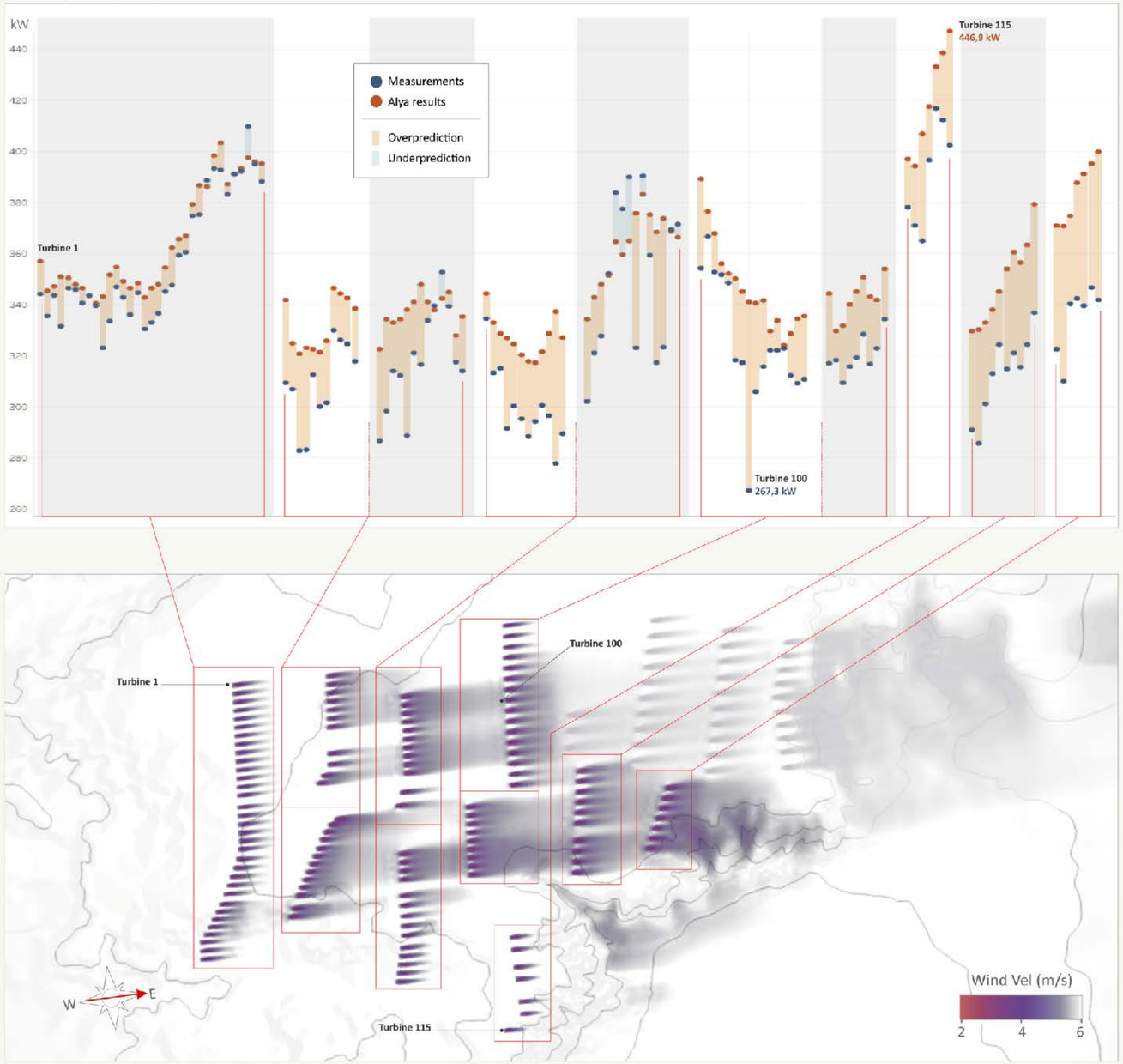


## Caveats

- Unclear if the windmill order is understandable in both charts- User tests needed
- Attenuated windmills excluded from the study can be confused with high speed values in the scale. New simulation needed
- More technical details (dates of the measurements, simulation method, model name, etc) needed for a stand-alone poster

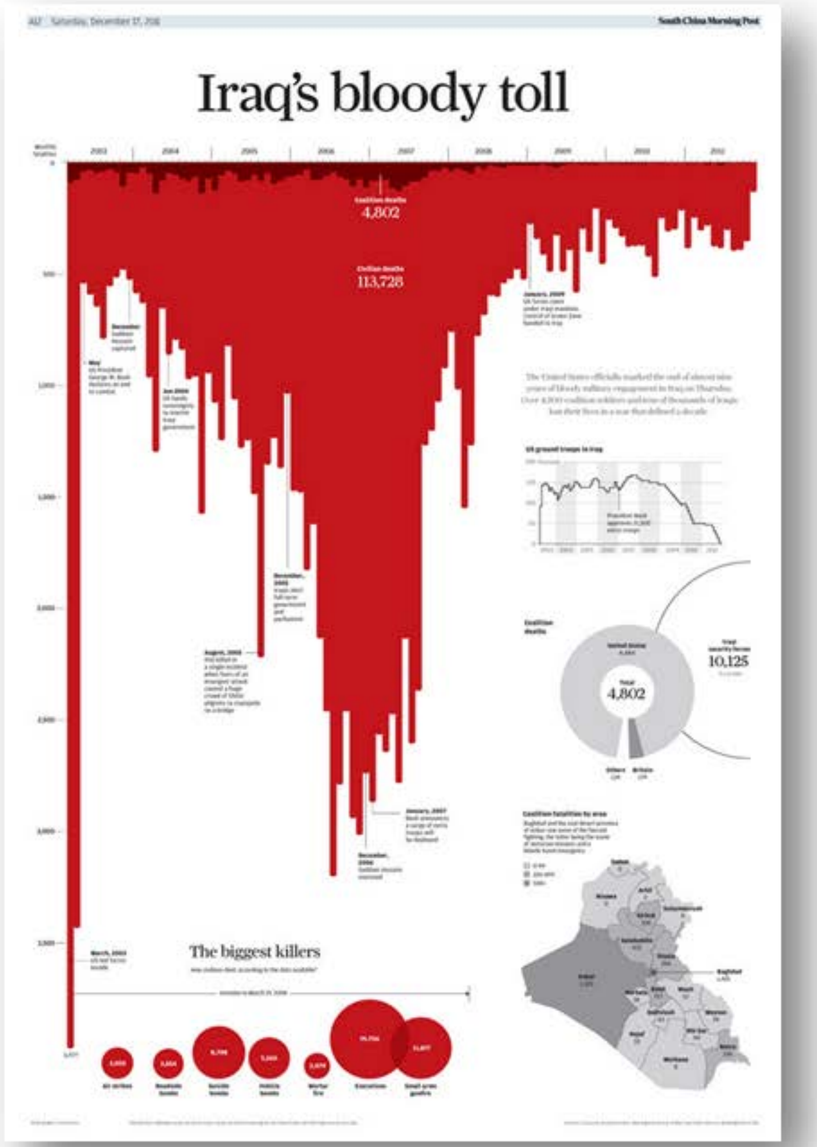
# Annual energy production

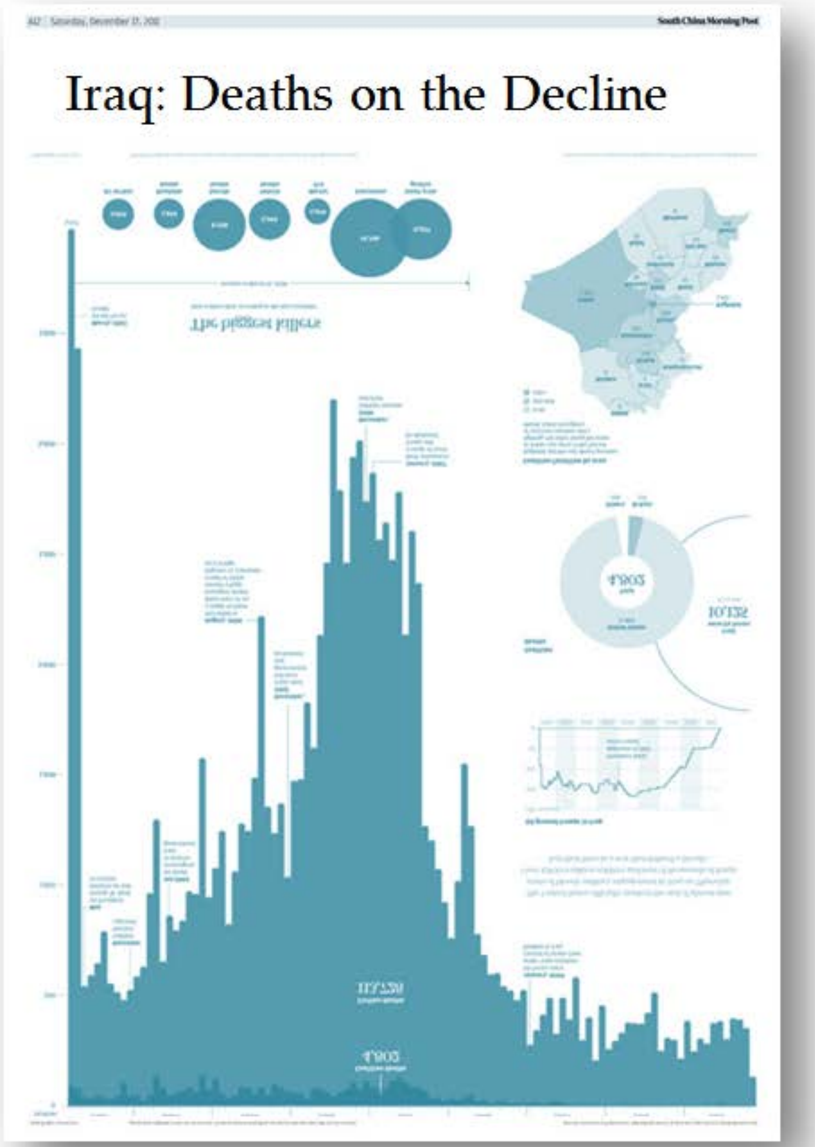
Estimation Vs. Measurement




Sisante windfarm

# There is always a story



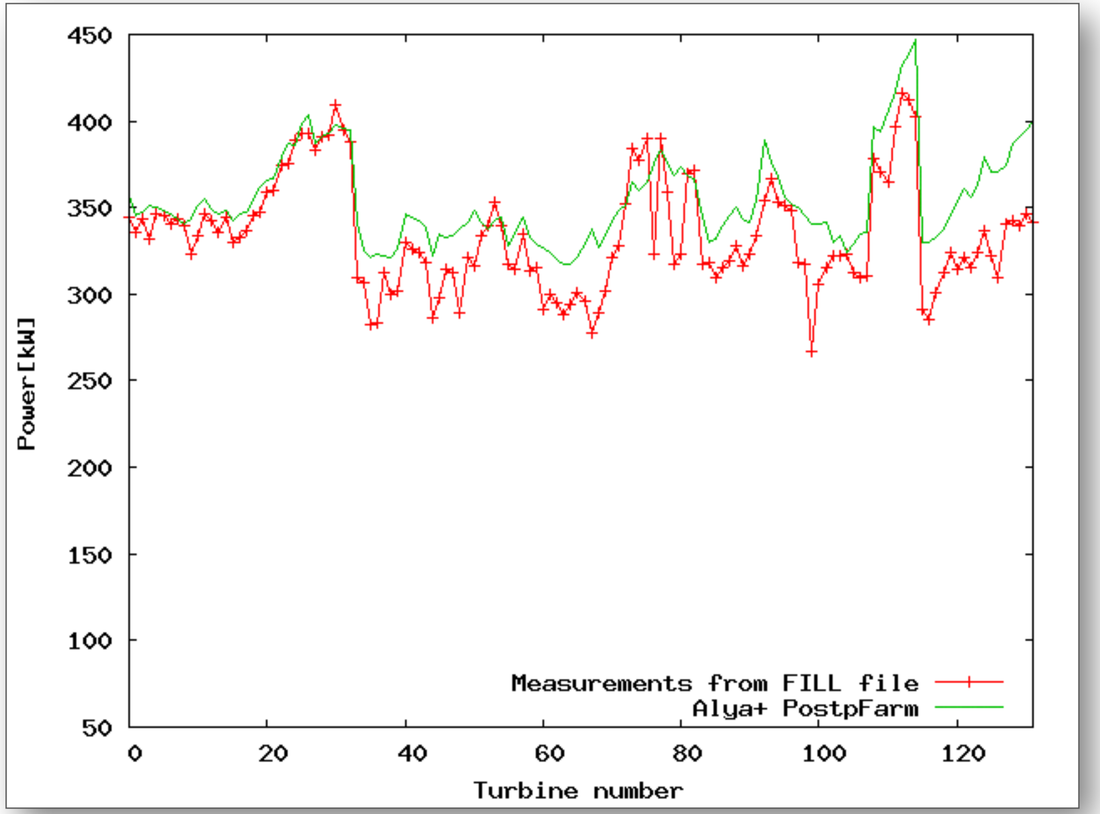

Andy Coatgrave 2014

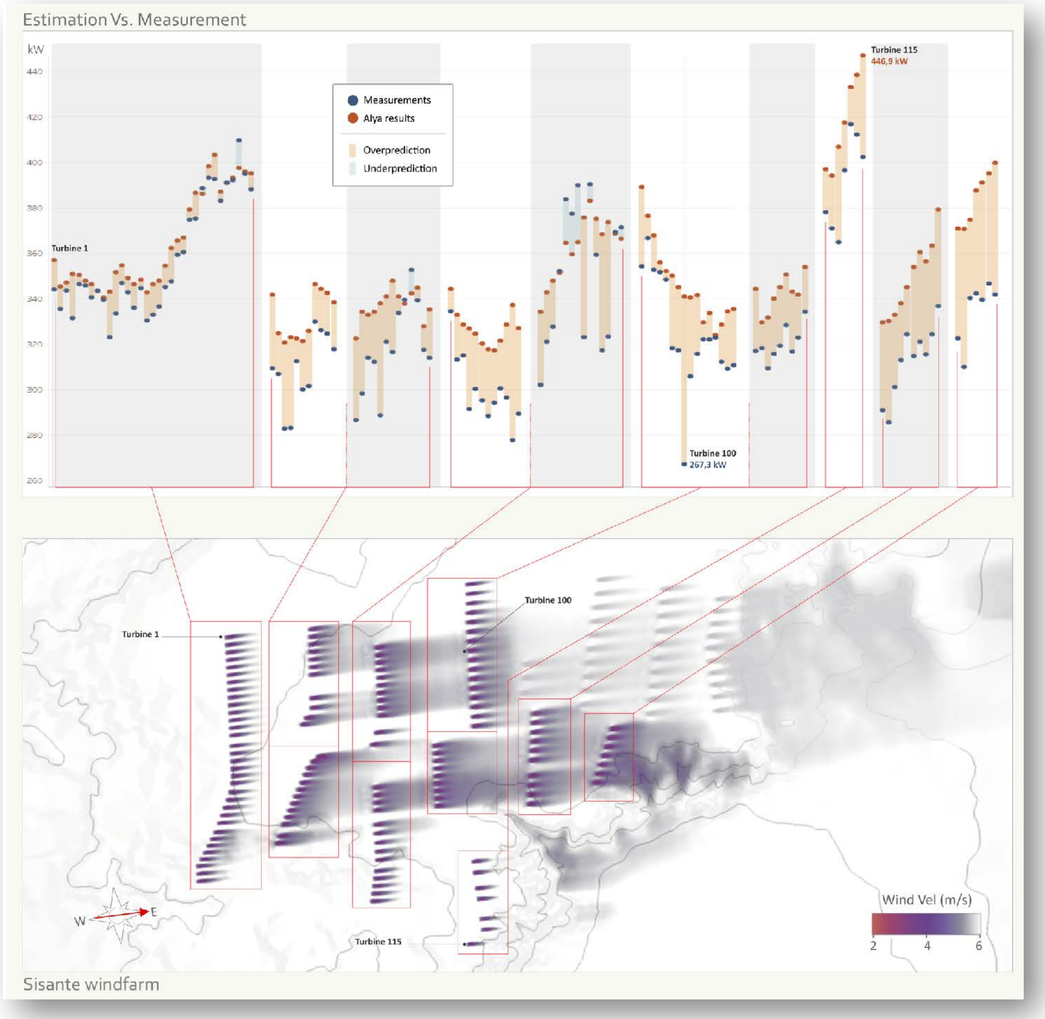
# There is always a story

My results somehow correlate to measurements

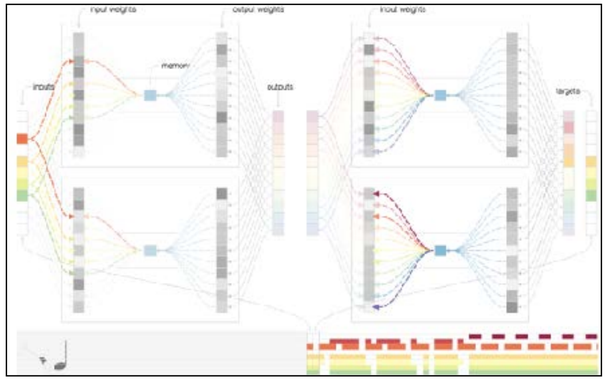
**Vs.**

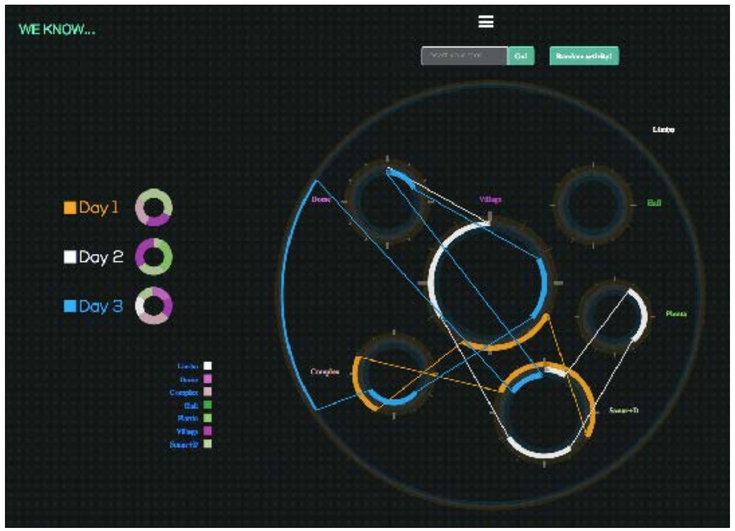
My model is not capturing turbulent wakes accurately

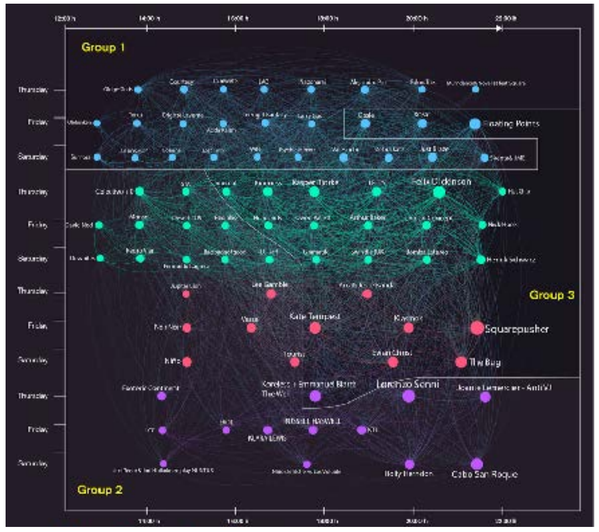


# BSC Data Analytics and Visualization Group



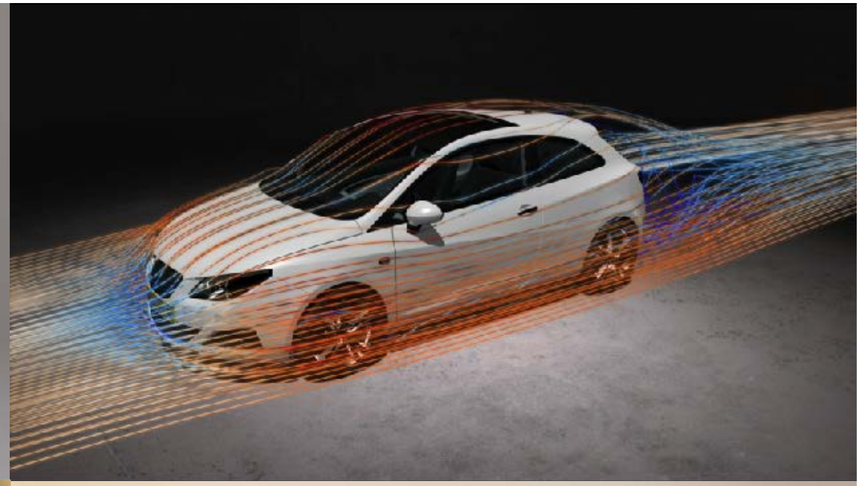



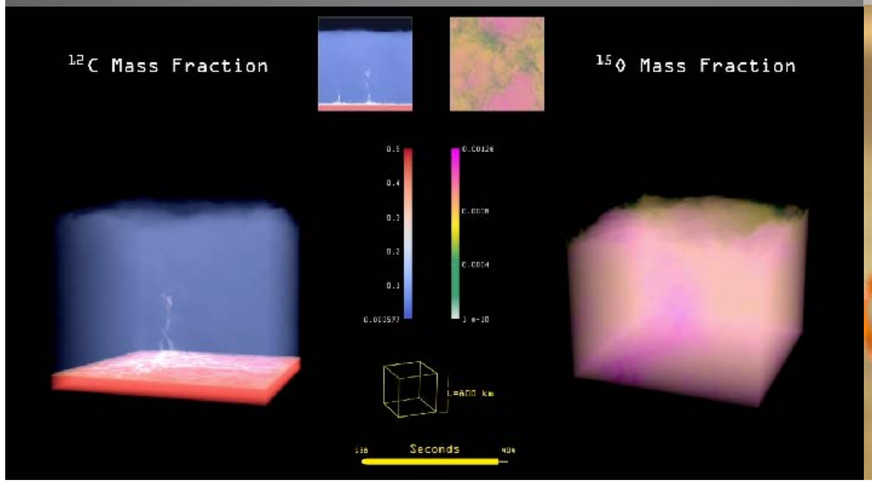



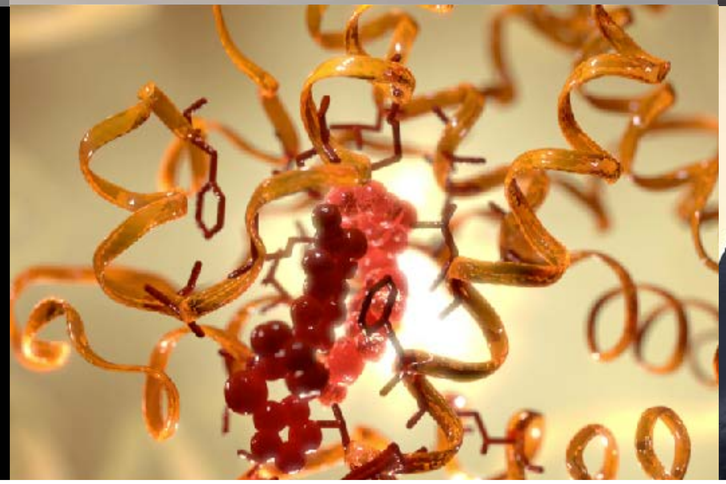

Hyper-realistic
data visualisation

# High-quality images under film industry standards

www.bsc.es/viz





# Over Produced data visualisation

Why?

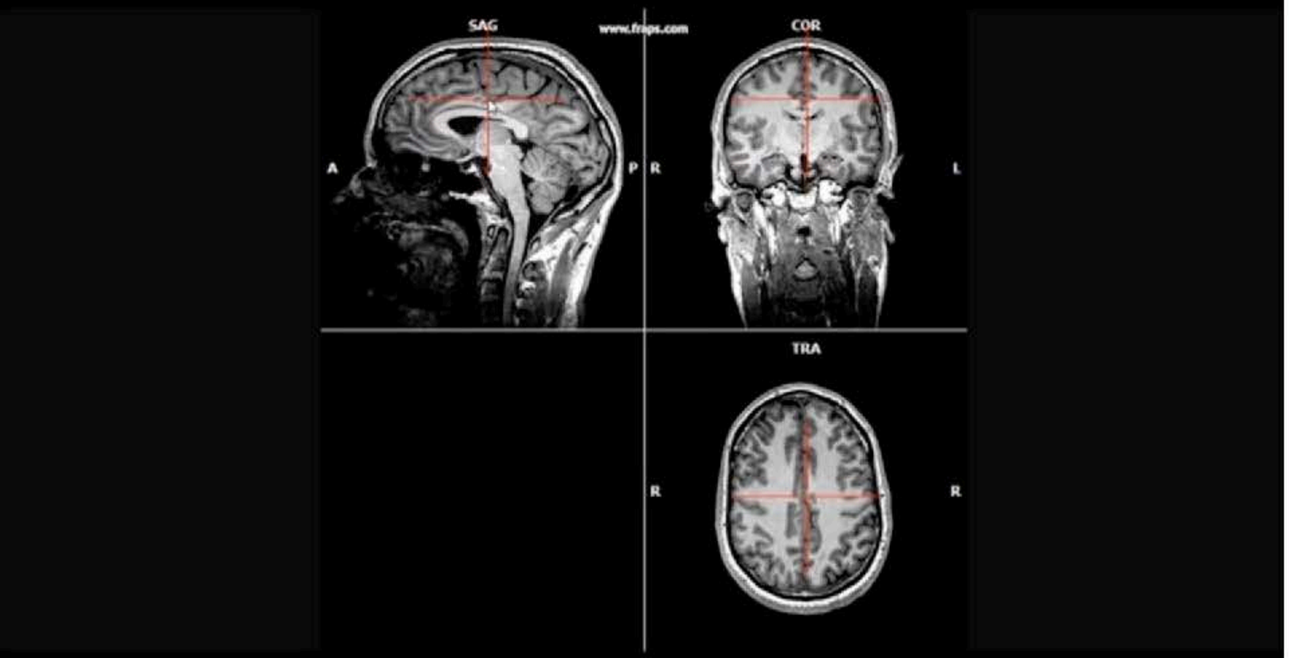
# High-end renders of data visualisation

Why?

- Highest possible impact
- Memorability

8th ANNUAL imagine SCIENCE Film Festival

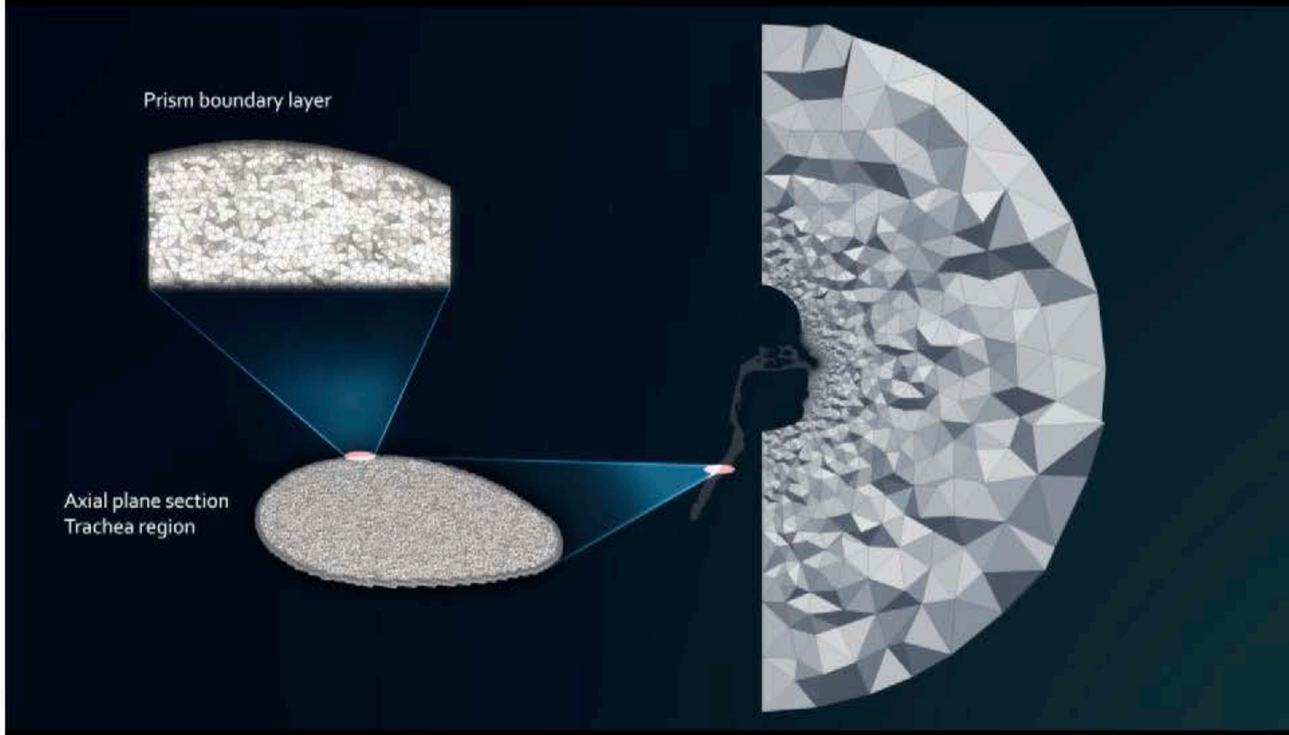
# Sniff




https://youtu.be/s6hDjPkhrPo

# Visualization of Airflow through the Human Respiratory System

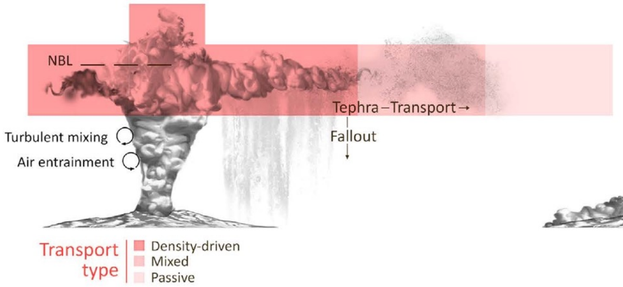
SC15 Austin, TX | hpc transforms

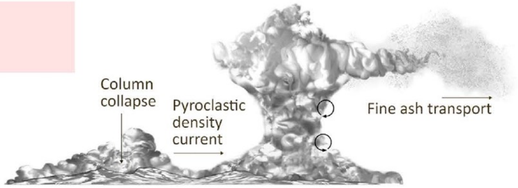


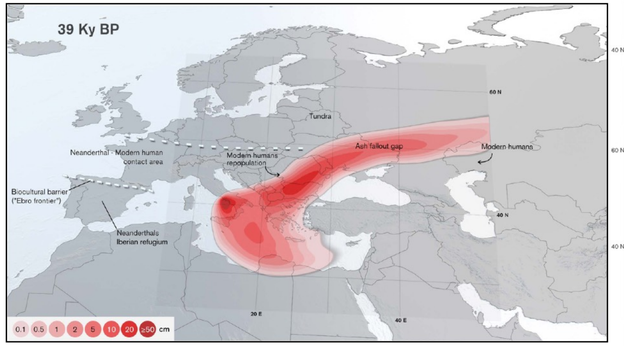

https://youtu.be/72OP03QOpXw

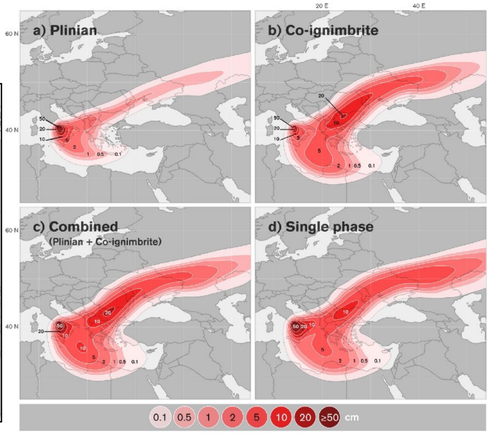
# Scientists

Scientific Reports **6**, Article number: 21220 (2016)



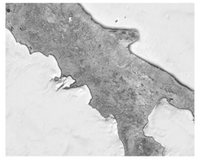

# General public

La Vanguardia (Spanish newspaper) 17/02/2016

Reconstrucción de la

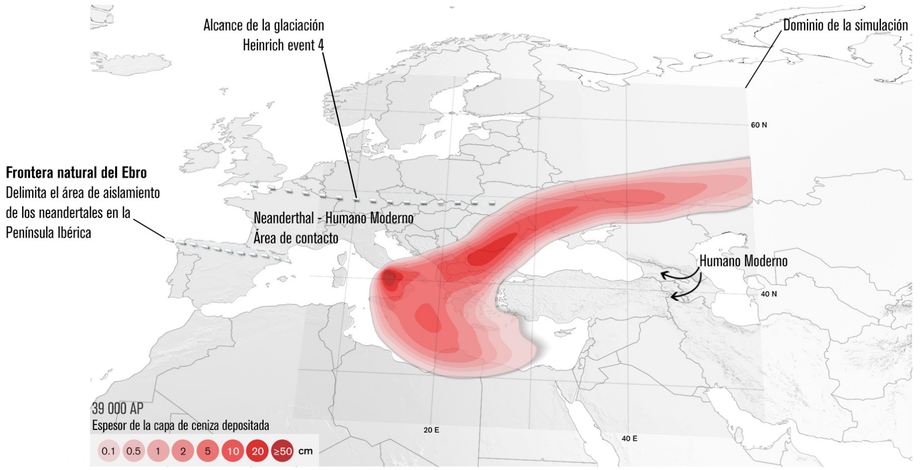
## SUPER-ERUPCIÓN
## Ignimbrita Campana

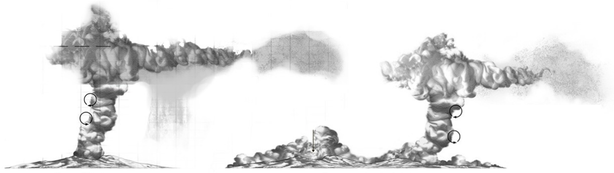
La ceniza expulsada cubrió más de 3 millones de Km²



La caldera ocupó lo que actualmente es el Golfo de Nápoles

**Impacto de la erupción en la transición del Paleolítico Medio al Paleolítico Superior**



Fase Pliniana

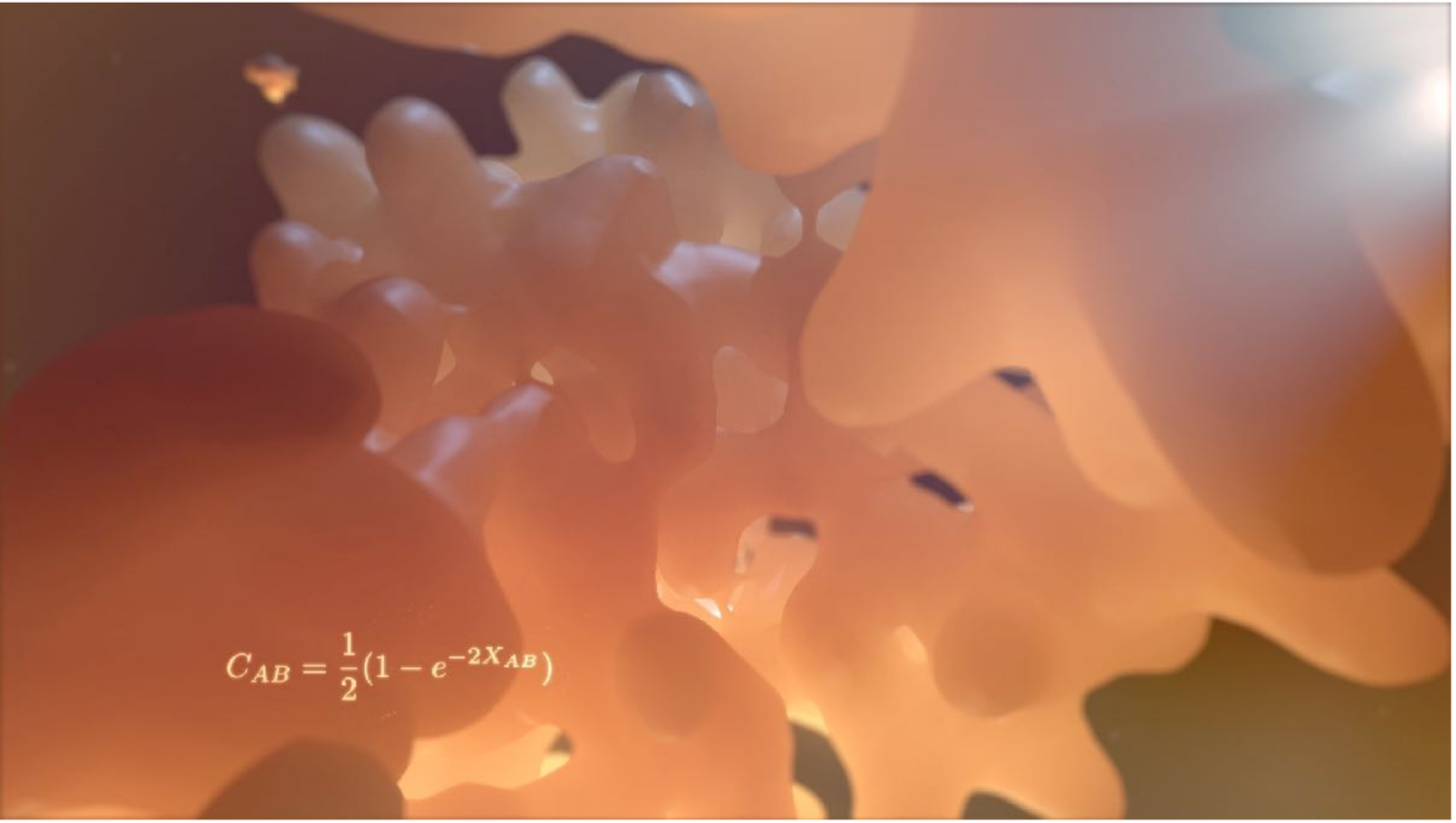
Fase co-ignimbrítica

La erupción tuvo dos fases: la Pliniana derivada de la erupción y la co-ignimbrítica causada por el co-lapso de la primera columna y los flujos piroclásti-cos. Las simulaciones computacionales nos ayudan a comprender mejor este tipo de fenóme-nos, analizando el rol de cada fase y sus conse-cuencias.

© BSC Visualization Team

BSC Viz Team

# Photo realistic data visualisation

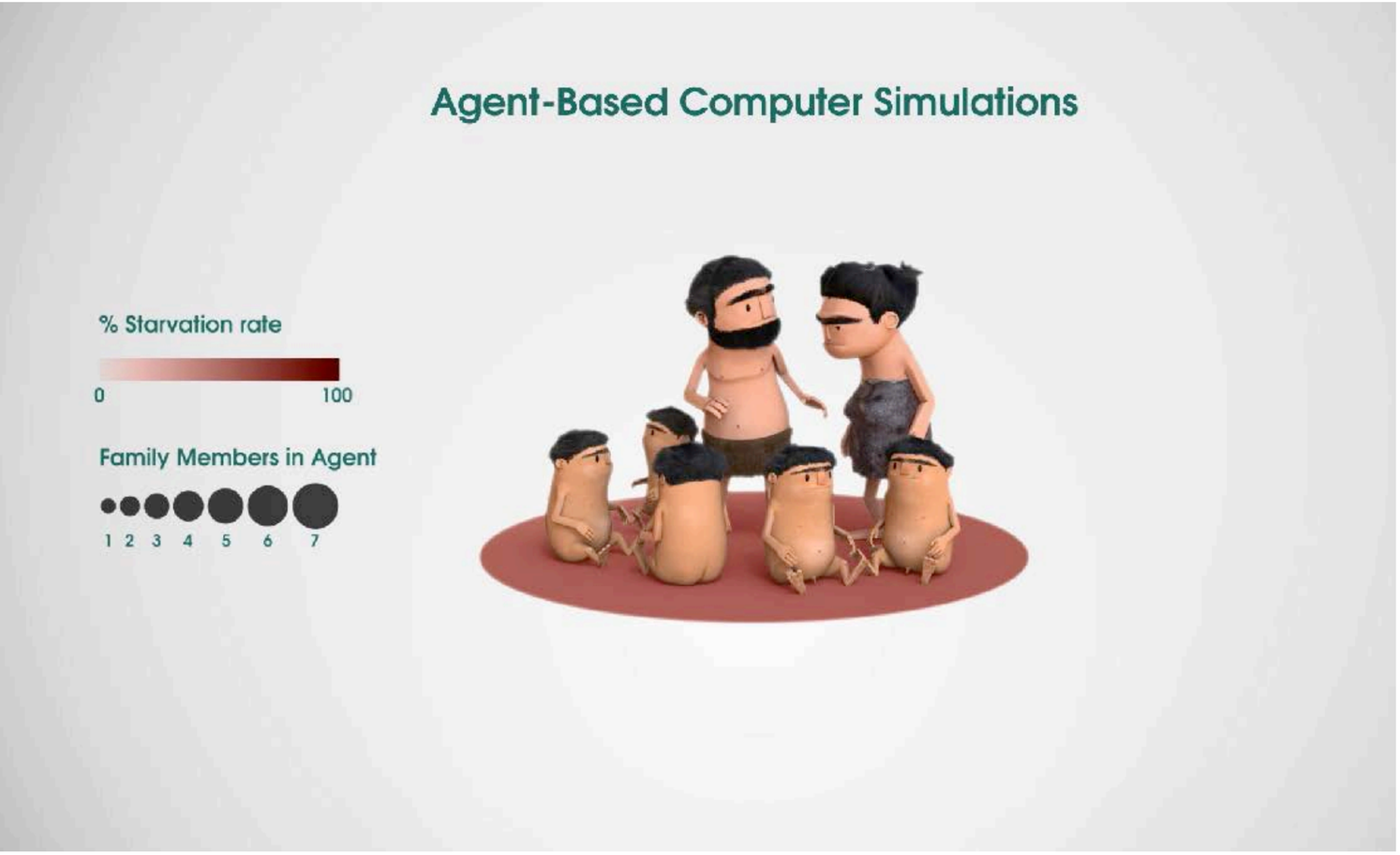



## How?

To achieve it, we need artist level of control over camera, light, animation, textures, and render quality

# Super nice data visualisation

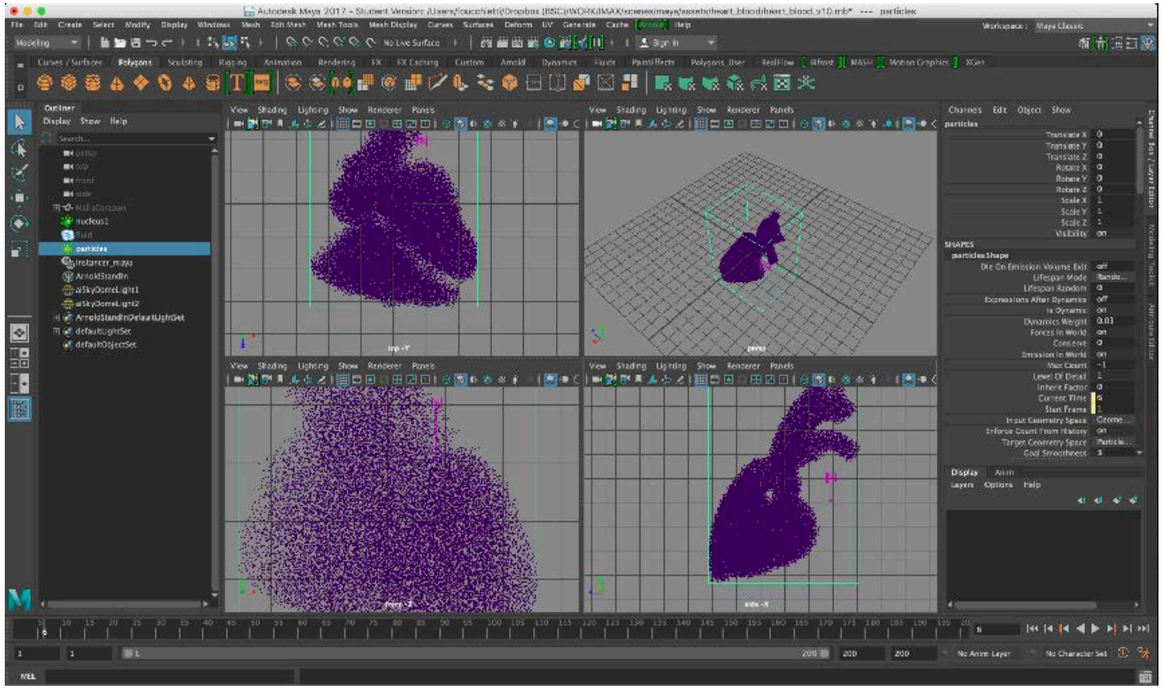
https://youtu.be/VooETfsDErM

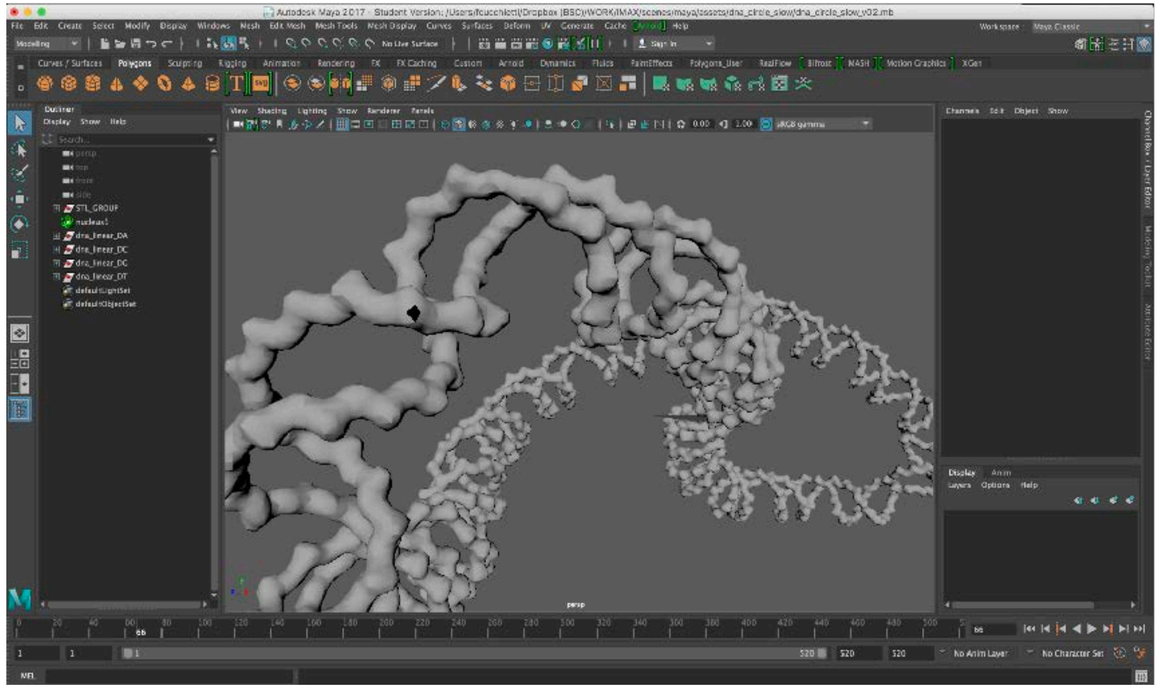



How?

Film industry tools are amazing

Film industry **people** too

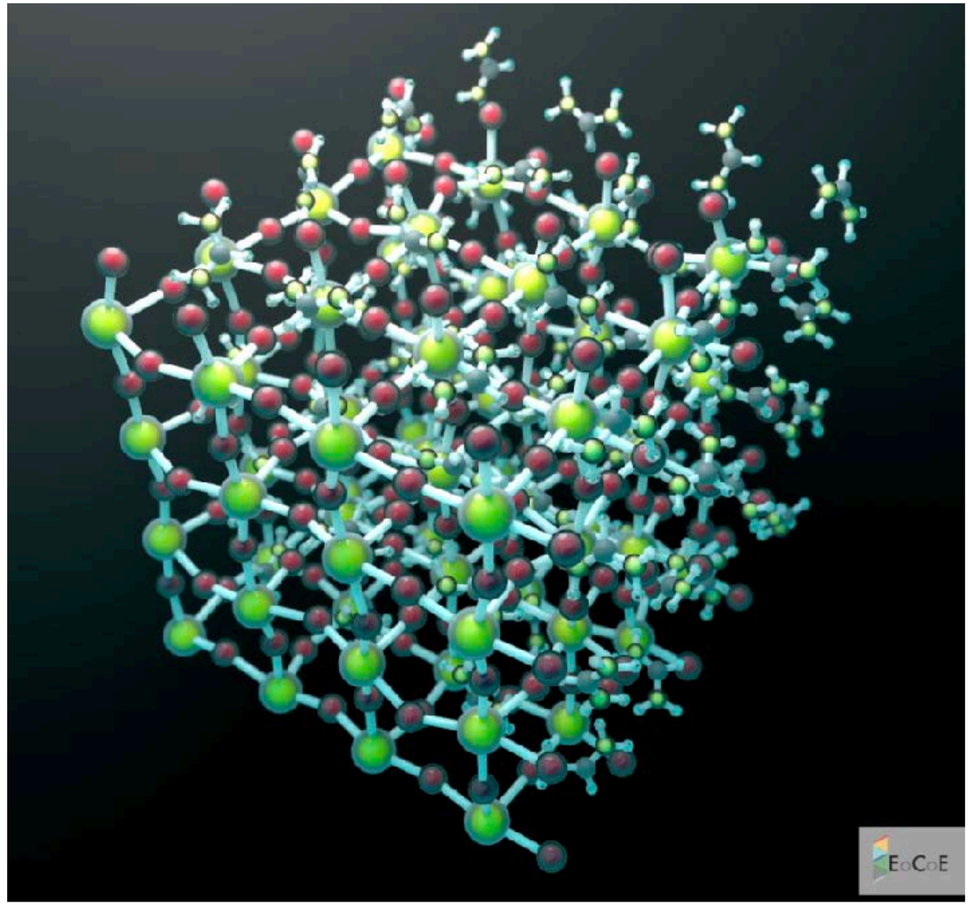




# Beautiful AND accurate

- Have scientists and artists work together
- Convert data from scientific software/ format into animation industry standards

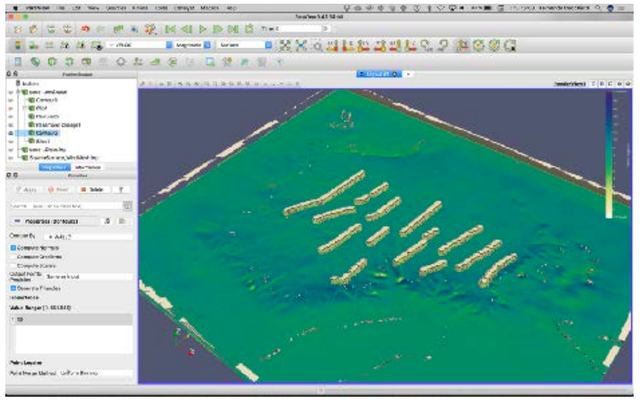
# Hyper cool
# data visualisation

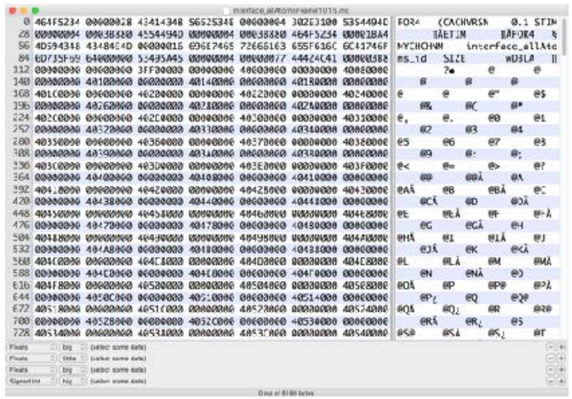


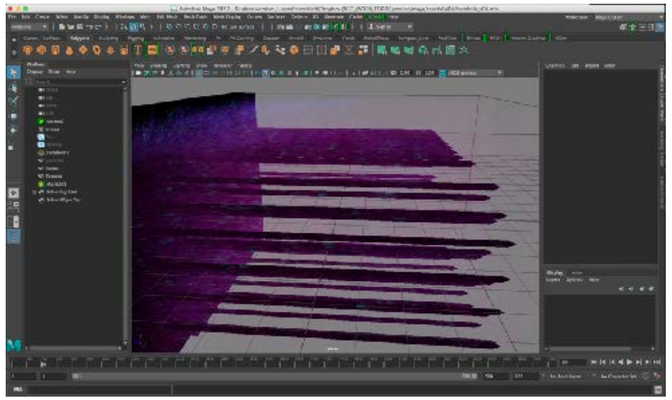

- Data conversion Tools and plugins
- Leverage standard formats: netcdf, vtk, ensi, geo
- Convert into:
  - volumetric data: Maya cache, OpenVDB
  - point/vector data: Maya cache, Partio
  - surface data: STL, OBJ, FBX, Alembic

# Typical pipeline

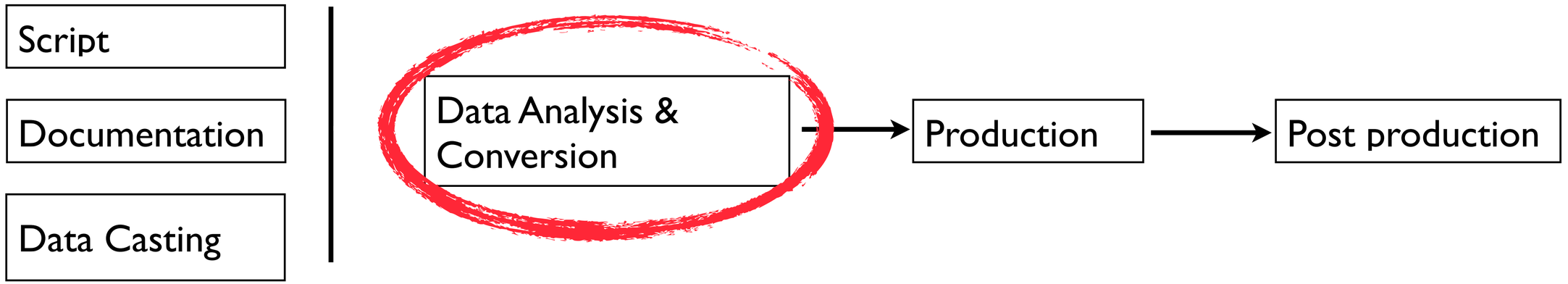
Load data in Sci-Viz software → Data forensics if necessary → Data Conversion → Fun in Maya, Blender, etc



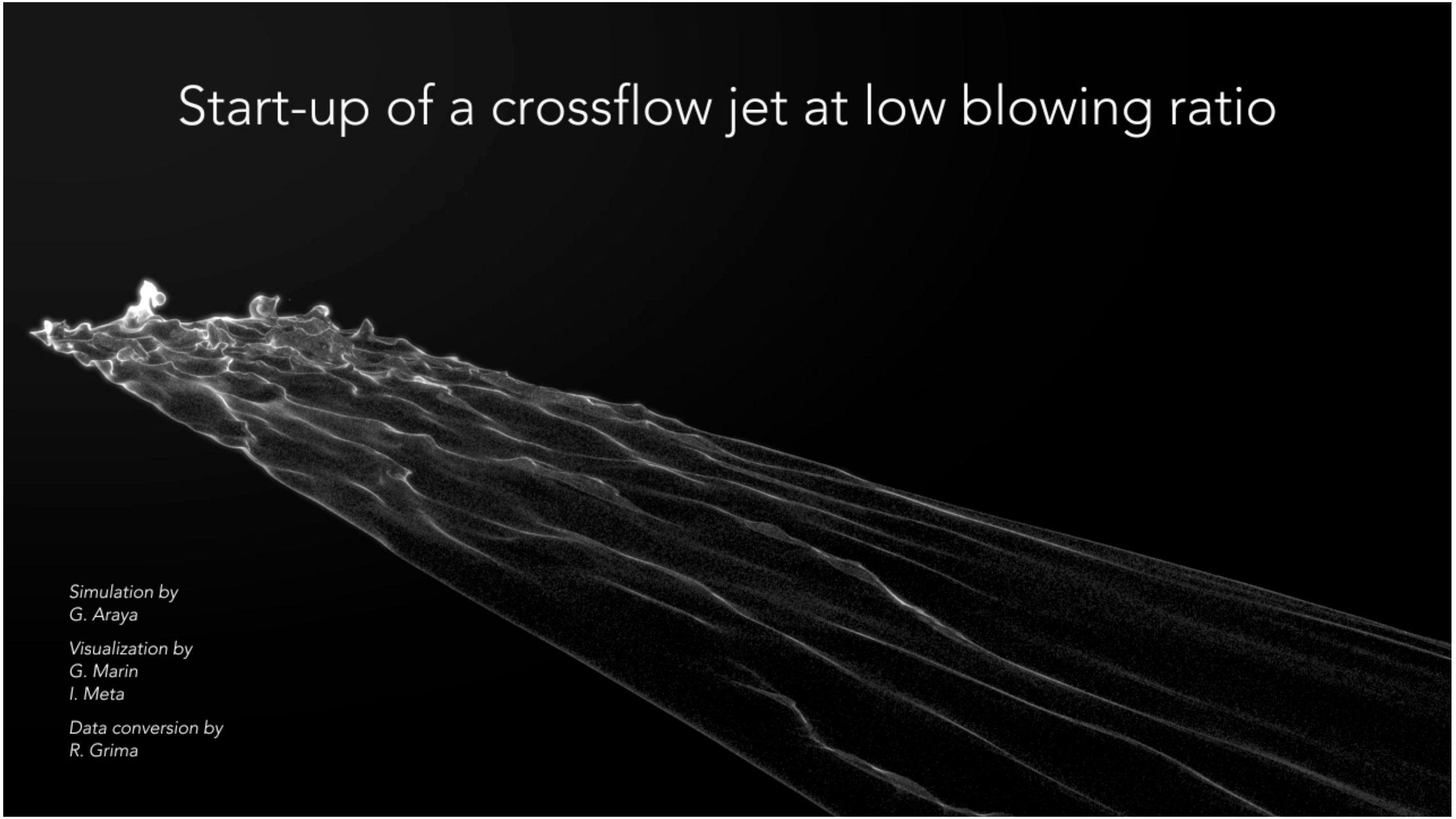




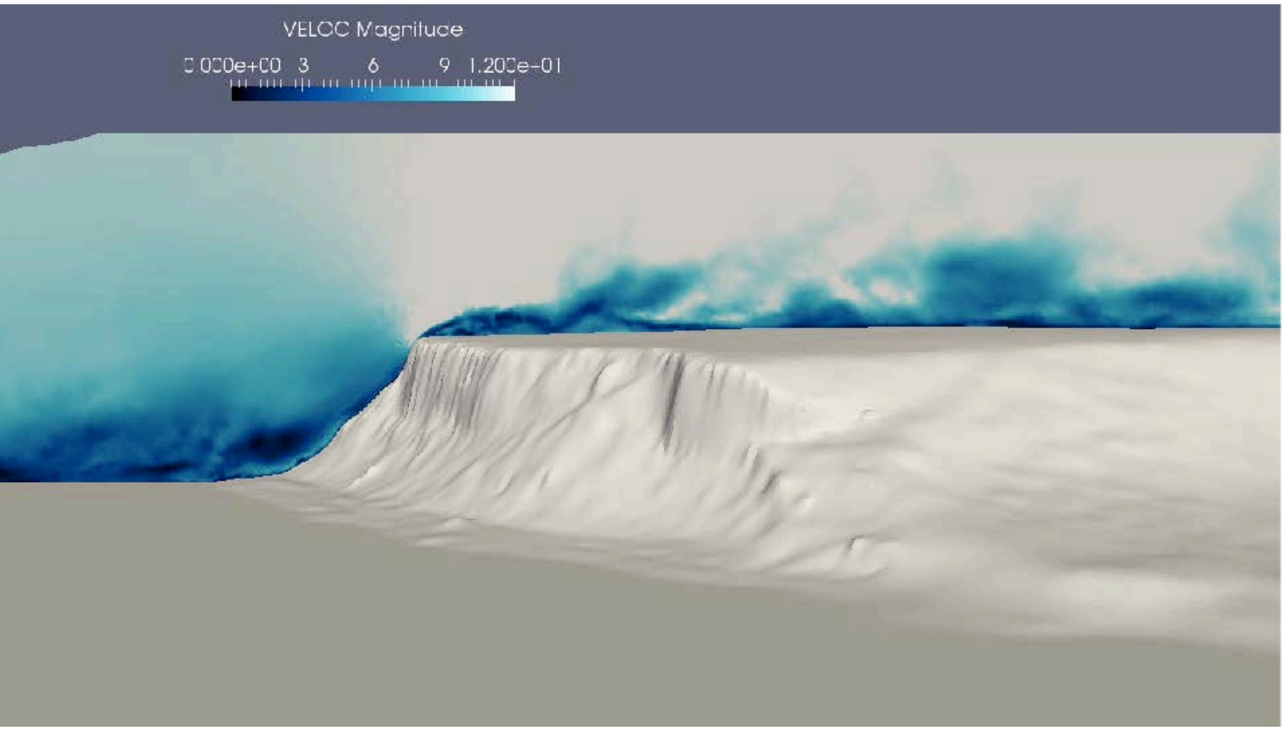
# Production pipeline

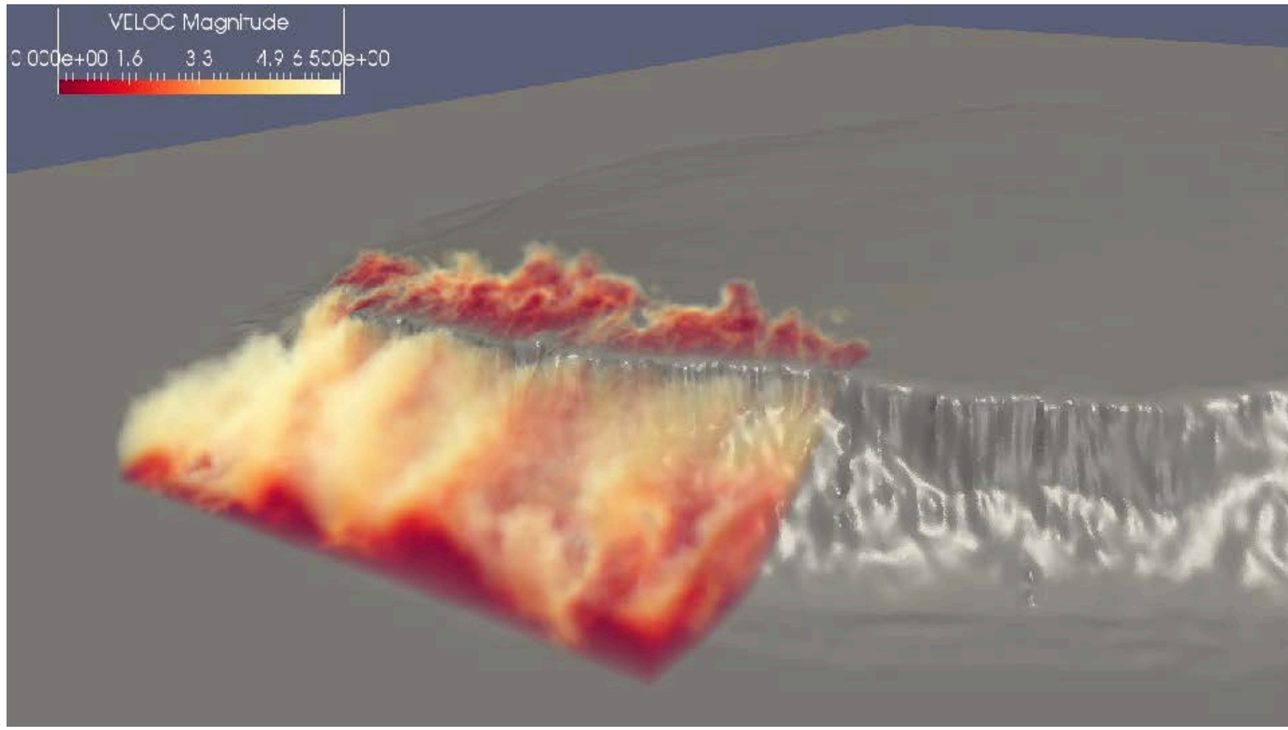


# Example CFD


Start-up of a crossflow jet at low blowing ratio
Simulation by
G. Araya
Visualization by
G. Marin
I. Meta
Data conversion by
R. Grima

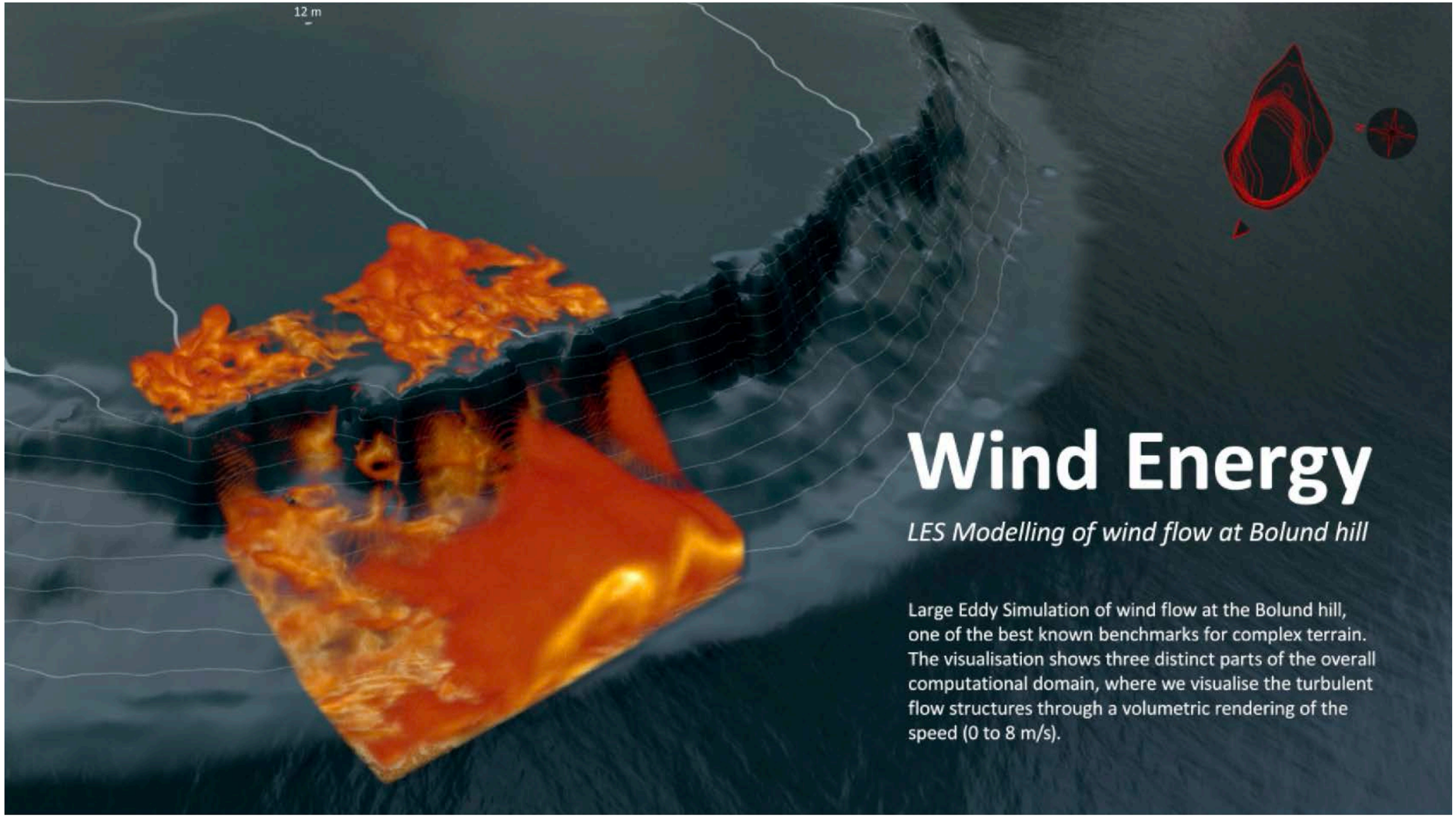
# BOLUND - Paraview



# Bolund - Maya

https://youtu.be/6qjG6tnjlXw



Atmospheric motion and water vapor
Feasability study with convection-permitting resolution
3km, 124.16M grid points
JUQUEEN@JSC, 4096 MPI tasks
Speed (m/s)
1
33
Height (km)
12
0.1
https://youtu.be/RIboJn8N130

Tools
are
Tools

# Software

## Data analysis / generation

R
Matlab
SPSS
etc

## DataViz

Tableau
Ncview
VMD
Paraview
R
Matlab
D3.js

## Diseño

Adobe Illustrator
Adobe Suite
Inkscape
Gimp
Blender

Steal it!

# Readings

## Books

- **Visual Strategies. Felice C. Frankel & Angela H. DePace, 2012**
- Visual Explanations. Edward Tufte, 1997
- **The Functional Art . Alberto Cairo, 2012**
- The truthful art. Alberto Cairo, 2016
- Designing with the Mind in Mind. Jeff Johnson, 2010
- Visual Thinking for Design. Colin Ware, 2008
- Information Visualization. Colin Ware, 2013
- Visualization Analysis & Design. Tamara Munzner, 2014

## Some articles

- https://kosara.net/publications.html
- http://ccom.unh.edu/vislab/colin_ware_pubs.html

## Websites

Color

- http://www.visualisingdata.com
- Subtleties of color
- https://visual.ly/blog/rainbow-color-scales
- The right color palettes

Inspiration

- http://www.thefunctionalart.com/
- https://eagereyes.org/
- http://www.informationisbeautiful.net/
- http://coolinfographics.com/
- http://flowingdata.com/
- http://fivethirtyeight.com/
- http://truth-and-beauty.net/
- http://www.gapminder.org/videos/

Examples of bad visualisations

- http://viz.wtf/